Structural modelling and robustness analysis of complex metabolic networks and signal transduction cascades

Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Biologisch-Pharmazeutischen Fakultät
der Friedrich-Schiller-Universität Jena

von Diplom-Chemiker Jörn Arnold Behre
geboren am 28. Juli 1973 **in** Bergisch Gladbach, Deutschland

Gutachter:

1. Prof. Dr. Stefan Schuster (Universität Jena)
2. PD Dr. Peter Dittrich (Universität Jena)
3. Prof. Dr. Ina Koch (Universität Frankfurt am Main)

Tag der öffentlichen Verteidigung: 23.03.2012

Die vorliegende Arbeit wurde am Lehrstuhl für Bioinformatik der Friedrich-Schiller-Universität Jena unter der Leitung von Prof. Dr. Stefan Schuster angefertigt.

In liebevollem Gedenken an meine Mutter

Gabriele Behre, geb. Pießenecker

⋆ 19. Oktober 1941 † 06. November 2010

Abstract

Due to the vast progresses in experimental methods and computational capacities molecular cell biology has faced a revolution in the last century. Whereas in the beginning of the 20th century there was nearly no knowledge about enzyme-catalysed reactions, structures of macromolecules, and information encoding in DNA and RNA, the enormous progress in genomics has pushed the field of molecular biology into the area of systems biology. Scientists are now able to quantify biological processes producing terabytes of data. But since data are not the same as knowledge, the need for mathematical approaches to evaluate these data has grown dramatically. In parallel, also the computational capacities to model biological processes have been expanded enormously and thus systems biology now contributes to holistic views on cellular processes, organs, and organisms. Present-day systems biology has two origins, on one side the more familiar "biological root" comprising, for instance, the discovery of DNA and its structure in the 1950s, and on the other side the less known "systemical root" covering, amongst others, non-equilibrium thermodynamics, feedback regulations in metabolism, and metabolic control analysis. Within the field of analysis of biological networks structural methods are particularly successful since they abstract from kinetic laws of the underlying reactions (which are very often unknown) as well as from the precise structure of metabolites.

The work I present in my PhD thesis contributes to this field of structural analysis of biological networks and in particular to the analysis of structural robustness of biological systems.

In the first part of my thesis, I introduce a new concept to calculate the structural robustness of metabolic networks based on the framework of elementary flux modes. I show that the number of elementary flux modes

itself is not an appropriate measure of structural robustness. The introduced robustness measures are based on the percentages of elementary flux modes remaining after knockouts of enzymes. After discussing the relevance of these measures with the help of simple examples, I demonstrate quantitatively that the metabolism of *Escherichia coli*, which must be able to adapt to varying conditions, is more robust than the metabolism of the human erythrocyte, which lives under much more homeostatic conditions.

In the second part of my thesis, I present a generalisation of this framework of structural robustness. In contrast to the previous study that is based on single knockouts, it is now possible to take also double and multiple knockouts into account. Thereby, the basic approach remains constant, just that knockout combinations are used instead of single reactions. I apply this extended framework to the amino acid anabolisms in *Escherichia coli* and human hepatocytes, and to the central metabolism in human erythrocytes. In this context, a comparison of the structural robustness of the human amino acid anabolisms with the one of *Escherichia coli* reveals that those amino acids that are essential for humans are also in *Escherichia coli* synthesised in a less robust manner than the non-essential amino acids. This result supports the hypothesis that in human the ability to synthesise these essential amino acids got lost during evolution (when this lacking was no longer lethal due to carnal nutrition) also due to structural weaknesses in metabolism.

In the third part of my thesis, I expand the concept of elementary flux modes in order to detect routes in signal-transduction networks consisting of enzyme cascades operating, for instance, by phosphorylation and dephosphorylation. The approach is based on the two ideas that the signal flow along each route through a network may not be diminished and that the system has to return to its original state after each signalling event. After illustrating this extended concept by several simple prototypic cascades it is applied to an example from insulin signalling.

In the outlook of my thesis, I discuss possibilities to extend my approach of quantifying structural robustness of metabolic networks also to other biological networks, such as signal-transduction or gene-regulatory networks.

Zusammenfassung

Aufgrund der enormen Fortschritte in Bezug auf die experimentellen Methoden und die Rechenkapazitäten moderner Computer hat die Molekularbiologie im letzten Jahrhundert eine Revolution erlebt. Während noch zu Beginn des 20. Jahrhunderts kaum Wissen über enzymatische Reaktionen, Strukturen von Makromolekülen oder die Kodierung von Informationen in DNA und RNA vorhanden war, hat der gewaltige Fortschritt in der Genetik die Molekularbiologie in Richtung Systembiologie vorangetrieben. Wissenschaftler sind inzwischen in der Lage, biologische Prozesse quantitativ zu messen und dabei Terabytes an Daten zu produzieren. Da aber Daten allein noch kein Wissen darstellen, ist der Bedarf an mathematischen Methoden zur Bewertung dieser Daten drastisch angestiegen. Parallel dazu sind aber auch die verfügbaren Rechnerressourcen zur Modellierung biologischer Prozesse stark angewachsen. Daher trägt die Systembiologie heute dazu bei, zelluläre Prozesse, Organe bzw. Organismen ganzheitlich zu betrachten. Die moderne Systembiologie hat zwei Wurzeln, zum einen den bekannteren „biologischen Ursprung“, der zum Beispiel die Entdeckung der DNA und ihrer Struktur in den 1950er Jahren beinhaltet, und zum anderen den „systemischen Ursprung“, der unter anderem die Nichtgleichgewichts-Thermodynamik, die Rückkopplungsregulationen in Stoffwechselsystemen oder die metabolische Kontrollanalyse umfasst. Innerhalb des Gebiets der Analyse biologischer Netzwerke sind strukturelle Methoden besonders erfolgreich, da sie zum einen von den Kinetiken der zugrundeliegenden Reaktionen (die häufig unbekannt sind) und zum anderen von der genauen inneren Struktur der beteiligten Metabolite abstrahieren.

Die hier in meiner Dissertation vorgestellte Arbeit leistet einen Beitrag zu strukturellen Analysemethoden biologischer Netzwerke, speziell zur Analyse

der strukturellen Robustheit biologischer Systeme.

Im ersten Teil meiner Dissertation stelle ich eine neue Methode vor, um die strukturelle Robustheit metabolischer Netzwerke basierend auf dem Konzept der elementaren Flussmoden zu bestimmen. Ich zeige, dass die Zahl der elementaren Flussmoden allein kein geeignetes Maß für die strukturelle Robustheit ist. Die vorgestellten Robustheitsmaße basieren auf den Anteilen an elementaren Flussmoden, die nach den Knockouts von Enzymen übrig bleiben. Nach der Diskussion der Relevanz dieser neuen Robustheitsmaße anhand von einfachen Beispielen zeige ich quantitativ, dass der Stoffwechsel des Bakteriums *Escherichia coli*, das in der Lage sein muss, sich an wechselnde Bedingungen anzupassen, robuster ist, als der Stoffwechsel des menschlichen Erythrozyten, der unter sehr gleichbleibenden Bedingungen lebt.

Im zweiten Teil meiner Dissertation stelle ich eine verallgemeinerte Version meines Robustheitskonzepts vor. Im Gegensatz zur vorangegangenen Studie, die nur auf Einfach-Knockouts basiert, ist es nun möglich, auch Doppel- und Vielfach-Knockouts zu betrachten. Der grundlegende Ansatz bleibt dabei gleich, es werden nun lediglich Knockout-Kombinationen anstelle von Einfach-Knockouts betrachtet. Ich wende dieses erweiterte Konzept auf den Aminosäurestoffwechsel von *Escherichia coli* und dem menschlichen Hepatozyten und außerdem auf den Zentralstoffwechsel des menschlichen Erythrozyten an. In diesem Zusammenhang zeigt ein Vergleich der Aminosäureanabolismen von *Escherichia coli* und dem menschlichen Hepatozyten, dass diejenigen Aminosäuren, die für den Menschen essenziell sind, auch in *Escherichia coli* weniger robust synthetisiert werden, als die nicht-essenziellen Aminosäuren. Dieses Ergebnis bestärkt die Hypothese, dass die menschliche Fähigkeit zur Synthese der nun essenziellen Aminosäuren im Verlauf der Evolution auch aufgrund von strukturellen Schwächen im Stoffwechsel verloren gegangen ist (als diese Verluste aufgrund von fleischlicher Ernährung kompensiert werden konnten).

Im dritten Teil meiner Dissertation verallgemeinere ich das Konzept der elementaren Flussmoden dahingehend, dass nun auch Wege durch Signaltransduktionsnetzwerke gefunden werden können, die aus Signalkaskaden aufgebaut sind. Solche Signalkaskaden können beispielsweise auf der Basis von Phosphorylierungen und Dephosphorylierungen funktionieren. Der An-

satz basiert auf den beiden Ideen, dass der Signalfluss entlang der Wege durch ein Netzwerk nicht vermindert werden darf, und dass sich das System zwischen zwei Signalflüssen regenerieren muss. Nach der Veranschaulichung dieses erweiterten Konzepts wende ich es auf ein Beispiel aus dem Insulin-Signalsystem an.

Abschließend diskutiere ich im Ausblick am Ende meiner Dissertation die Möglichkeiten, meinen Ansatz der Bestimmung von strukturellen Robustheiten metabolischer Netzwerke auch auf andere biologische Netzwerke, wie zum Beispiel Signaltransduktionsnetzwerke oder genregulatorische Netzwerke auszudehnen.

Danksagung

So sehr ich es auch bedauere, aber meine Zeit als Doktorand am Lehrstuhl für Bioinformatik an der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität in Jena neigt sich nun ihrem Ende zu. Ich möchte daher meinen besonderen Dank an den gesamten Lehrstuhl richten. Es ist schon ein echtes Glück, an einem Lehrstuhl mit so vielen, so netten und hilfsbereiten Kollegen promovieren zu dürfen. Ohne diese freundliche Atmosphäre hätte meine Dissertation in dieser Form nicht entstehen können. Besonders hervorheben möchte ich dabei die stets wohlwollende und sehr geduldige Unterstützung meines Betreuers Prof. Dr. Stefan Schuster.

Weiterhin möchte ich meinen Mitautoren und Kooperationspartnern danken: Dr. Thomas Wilhelm vom Institute of Food Research in Norwich, Großbritannien, Dr. Axel von Kamp, Dr. Steffen Klamt und Regina Samaga vom Max-Planck-Institut für Dynamik komplexer technischer Systeme, Prof. Dr. Dr. Eytan Ruppin von der Universität in Tel Aviv, Israel und Luís Filipe de Figueiredo, Dr. Thomas Hinze und Dr. Christoph Kaleta am Lehrstuhl für Bioinformatik.

Ein ganz großer Dank gebührt auch Frau Dr. Ina Weiß, die hier am Lehrstuhl für Bioinformatik die wissenschaftliche Informationsstelle der Biologisch-Pharmazeutischen Fakultät betreibt, und der keine Veröffentlichung verborgen bleibt, und sei sie noch so exotisch.

Ebenso will ich natürlich auch den Gutachtern meiner Dissertation, Prof. Dr. Stefan Schuster, Prof. Dr. Ina Koch und PD Dr. Peter Dittrich meinen herzlichen Dank aussprechen für ihre Mühen beim Durcharbeiten und Bewerten meiner Dissertation.

Außerdem danke ich an dieser Stelle allen, die mir mit Rat und Tat bei Fragen zur Seite gestanden haben, zum einen noch einmal ganz besonders Dr. Thomas Hinze, und zum anderen allen, die ich hier nicht aufzählen kann.

In ganz besonderem Maße aber möchte ich meinen Eltern danken, nicht nur dafür, dass ich ohne sie ja gar nicht existieren würde, sondern vor allem für ihre unerschöpfliche Geduld mit mir.

Contents

Abstract **v**

Zusammenfassung **vii**

Danksagung **xi**

Contents **xv**

Glossary **xvii**

1 General Introduction **1**

1.1 Motivation 1

1.2 Structural modelling 5

1.2.1 The concept of Elementary Flux Modes 5

1.2.2 Petri Nets 16

1.2.3 Chemical Organisation Theory 22

1.2.4 Flux Balance Analysis 24

1.2.5 Extreme Pathways 25

1.2.6 Boolean Networks 26

1.3 Robustness 32

1.4 Outline of the thesis 40

2 Structural robustness concerning single knockouts **43**

T. Wilhelm, J. Behre, and S. Schuster. Analysis of structural robustness of metabolic networks. *IEE Proceedings - Systems Biology*, 1(1):114–120, 2004. 44

3 Structural robustness concerning multiple knockouts **51**

J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, and S. Schuster. Structural robustness of metabolic networks with respect to multiple knockouts. *Journal of Theoretical Biology*, 252(3):433–441, 2008. 52

4 Modelling signal transduction in enzyme cascades **61**

J. Behre and S. Schuster. Modeling signal transduction in enzyme cascades with the concept of elementary flux modes. *Journal of Computational Biology*, 16(6):829–844, 2009. 62

5 General Discussion **79**

Bibliography **118**

Appendix **119**

A Supplementary Material to Chapter 3 119

A.1 Supplement 1 119

A.2 Supplement 2 124

A.3 Supplement 3 127

A.4 Supplement 4 129

A.5 Supplement 5 133

A.6 Supplement 6 134

B Supplementary Material to Chapter 4 135

B.1 Supplement 1a 135

B.2 Supplement 1b 135

B.3 Supplement 2a 137

B.4 Supplement 2b 137

B.5 Supplement 3a 139

B.6 Supplement 3b 140

B.7 Supplement 4a 142

B.8 Supplement 4b 142

B.9 Supplement 5a 144

B.10 Supplement 5b 145

B.11 Supplement 6a 147

B.12 Supplement 6b 147
B.13 Supplement 7a 149
B.14 Supplement 7b 149
B.15 Supplement 8a 151
B.16 Supplement 8b 151

Beitragende Autoren **155**

Über den Autor **157**
Lebenslauf 157
Publikationen 158
Vorträge 159
Posterpräsentationen 160
Lehre 161

Erklärung **163**

Glossary

$\mathbf{B}$	Incidence matrix (within the theories of Petri nets and Boolean networks), equivalent to the stoichiometric matrix $\mathbf{N}$ of a metabolic network
F_i	The fragility coefficient F_i, introduced by Klamt and Gilles (2004), is defined as the reciprocal of the average size of all minimal cut sets in which a reaction (enzyme) E_i participates.
F	The network (overall) fragility coefficient F, introduced by Klamt and Gilles (2004), is defined as the average of all fragility coefficients F_i over all reactions (enzymes) E_i.
$\mathbf{K}^{(\mathbf{l})}$	Left null-space matrix
$\mathbf{K}^{(\mathbf{r})}$	Right null-space matrix, also called kernel matrix
M	Flux mode
$\mathbf{N}$	Stoichiometric matrix
V	Flux vector
`INA`	The program `Integrated Net Analyzer` is a tool package for the analysis of Petri nets and Coloured Petri nets (see `http://www2.informatik.hu-berlin.de/lehrstuehle/automaten/ina/`).

`PED`	The program `PED` (see `http://www-dssz.informatik.tu-cottbus.de/index.html?/software/ped.html`) is a hierarchical Petri-net editor and the predecessor of the program `Snoopy`.
ADP	Adenosine 5'-diphosphate
AMP	Adenosine 5'-monophosphate
API	Application programming interface
ATP	Adenosine 5'-triphosphate
BN	Boolean network
cAMP	Adenosine 3',5'-cyclic phosphate, second messenger
Cardinality d	The cardinality d is the combinatorial depth of a knockout. Thus, a single knockout has $d = 1$, a double knockout $d = 2$, and so on.
CDP	Cytidine 5'-diphosphate
CMP	Cytidine 5'-monophosphate
CREB(PS)	Cyclic AMP-responsive element-binding protein (serine phosphorylated), transcription factor
CTP	Cytidine 5'-triphosphate
DIC_2	Succinate-fumarate antiport (between cytosol and mitochondria) in *Homo sapiens* (see Palmieri, 2004; Visser *et al.*, 2007; Palmieri, 2008)
EC 1.1.1.1	The Enzyme Commission number (EC number) is a numerical classification system for enzymes, maintained by the "Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)". The classification scheme is

	based on the chemical reactions that the enzymes catalyse. The EC number 1.1.1.1, taken here as an example, represents the enzyme "Alcohol:NAD+ oxidoreductase". For further details see http://www.chem.qmul.ac.uk/iubmb/enzyme/.
EFM	Elementary flux mode
Ery_up	Artificial uptake reaction for D-Erythrose 4-phosphate in the Metatool model of *Escherichia coli* amino acid anabolism
ESM	Elementary signalling mode (or elementary signalling pathway); since the signal transduction pathways that we calculate in enzyme cascades with our extended framework of elementary flux modes (see Behre and Schuster, 2009, Chapter 4) do no longer coincide with a real mass flow but with a seeming mass flow, one should no longer call them elementary flux modes.
FBA	Flux balance analysis
G3P_up	Artificial uptake reaction for D-Glyceraldehyde 3-phosphate in the Metatool model of *Escherichia coli* amino acid anabolism
GDP	Guanosine 5'-diphosphate
GLUT2 transporter	GLUT is a family of glucose carriers (or better glucose transport facilitators) that transport glucose in the cells by facilitated diffusion through the cellular membranes. GLUT2 is responsible for glucose transport in hepatocytes and pancreatic β-cells.
GLUT4 transporter	GLUT is a family of glucose carriers (or better glucose transport facilitators) that transport glucose in the cells by facilitated diffusion through the cellular

	membranes. GLUT4 is responsible for glucose transport in adipocytes and muscle cells. The number of GLUT4 transporters is increased by insulin stimulation.
GMP	Guanosine 5'-monophosphate
GTP	Guanosine 5'-triphosphate
HOT theory	*Highly optimized tolerance theory* (see Carlson and Doyle, 2002)
KEGG database	*KEGG: Kyoto Encyclopedia of Genes and Genomes*; comprehensive database comprising biological data from whole genomes up to metabolic network maps. The database is located at the Kyoto University Bioinformatics Center in Japan. It is maintained by Prof. Minoru Kanehisa.
LIH	Logical interaction hypergraph
LSSA	Logical steady-state analysis
MCA	Metabolic control analysis
MCS	Minimal cut set
MCT-set	Maximal Common Transition Set, equivalent (but less strictly defined) within the theory of Petri nets to a reaction subset in the concept of elementary flux modes
MIS	Minimal intervention set
NAD^+	Nicotinamide adenine dinucleotide, oxidised form
$NADH+H^+$	Nicotinamide adenine dinucleotide, reduced form
$NADP^+$	Nicotinamide adenine dinucleotide phosphate, oxidised form

NADPH+H^+	Nicotinamide adenine dinucleotide phosphate, reduced form
NDP	Nucleoside diphosphate, unspecific symbol for ADP, CDP, GDP, or UDP
NMP	Nucleoside monophosphate, unspecific symbol for AMP, CMP, GMP, or UMP
NTP	Nucleoside triphosphate, unspecific symbol for ATP, CTP, GTP, or UTP
ODE	Ordinary differential equation
OGC	Malate-oxoglutarate antiport (between cytosol and mitochondria) in *Homo sapiens* (see Palmieri, 2004; Visser *et al.*, 2007; Palmieri, 2008)
P-invariant	P-invariants in the theory of Petri nets correspond to mass conservation relations in the concept of elementary flux modes.
PDE	Partial differential equation
Petri net	Bipartite graph consisting of two types of nodes, places and transitions, connected by arcs
PIP_3	Phosphatidylinositol (3,4,5)-trisphosphate, second messenger
poxA	Also called yjeA or STM4344, gene in *Salmonella enterica* encoding lysyl-tRNA synthetase
Pro_tr	Proline-glutamate antiport (between cytosol and mitochondria) in *Homo sapiens* (see Atlante *et al.*, 1996; Porter, 2000)
PTS	Phosphotransferase system (cf. Papin and Palsson, 2004)

Pyr_up — Artificial uptake reaction for pyruvate in the `Metatool` model of *Homo sapiens* amino acid anabolism

R00209 — KEGG identifier for the reaction in *Homo sapiens* catalysed by the mitochondrial multienzyme pyruvate dehydrogenase complex consisting of the enzymes "Pyruvate:[dihydrolipoyllysine-residue acetyltransferase]-lipoyllysine 2-oxidoreductase (decarboxylating, acceptor-acetylating)" (EC 1.2.4.1), "Acetyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-acetyltransferase" (EC 2.3.1.12), and "Protein-N6-(dihydrolipoyl)lysine:NAD+ oxidoreductase" (EC 1.8.1.4)

R00258 — KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "L-alanine:2-oxoglutarate aminotransferase" (EC 2.6.1.2); the existence of this enzyme is not affirmed by the KEGG database but by the EcoCyc database

R00268 — KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "Isocitrate:NADP+ oxidoreductase (decarboxylating)" (EC 1.1.1.42)

R00342 — KEGG identifier for the reaction in *Homo sapiens* catalysed by the enzyme "(S)-malate:NAD+ oxidoreductase" (EC 1.1.1.37)

R00344 — KEGG identifier for the reaction in *Homo sapiens* catalysed by the enzyme "Pyruvate:carbon-dioxide ligase (ADP-forming)" (EC 6.4.1.1)

R00351 — KEGG identifier for the reaction in *Homo sapiens* catalysed by the enzyme "Acetyl-CoA:oxaloacetate C-acetyltransferase [thioester-hydrolysing, (pro-S)-carboxymethyl forming]" (EC 2.3.3.1)

R00355 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme “L-aspartate:2-oxoglutarate aminotransferase” (EC 2.6.1.1)

R00480 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme “ATP:L-aspartate 4-phosphotransferase” (EC 2.7.2.4)

R00582 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme “O-phosphoserine phosphohydrolase” (EC 3.1.3.3)

R00586 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme “Acetyl-CoA:L-serine O-acetyltransferase” (EC 2.3.1.30)

R00782 KEGG identifier for the reaction in *Homo sapiens* catalysed by the enzyme “L-cystathionine cysteine-lyase (deaminating; 2-oxobutanoate-forming)” (EC 4.4.1.1)

R00897 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme “O3-acetyl-L-serine:hydrogen-sulfide 2-amino-2-carboxyethyltransferase” (EC 2.5.1.47)

R01061 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme “D-glyceraldehyde-3-phosphate:NAD+ oxidoreductase (phosphorylating)” (EC 1.2.1.12)

R01082 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme “(S)-malate hydro-lyase (fumarate-forming)” (EC 4.2.1.2)

R01324 KEGG identifier for the reaction in *Escherichia coli* and *Homo sapiens* catalysed by the enzyme “Cit-

rate(isocitrate) hydro-lyase (cis-aconitate-forming)" (EC 4.2.1.3)

R01714 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "5-O-(1-carboxyvinyl)-3-phosphoshikimate phosphate-lyase (chorismate-forming)" (EC 4.2.3.5)

R01773 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "L-homoserine:NAD(P)+ oxidoreductase" (EC 1.1.1.3)

R01826 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "Phosphoenolpyruvate:D-erythrose-4-phosphate C-(1-carboxyvinyl)transferase (phosphate-hydrolysing, 2-carboxy-2-oxoethyl-forming)" (EC 2.5.1.54)

R01899 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "Isocitrate:NADP+ oxidoreductase (decarboxylating)" (EC 1.1.1.42)

R02164 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "Succinate:ubiquinone oxidoreductase" (EC 1.3.5.1)

R02291 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "L-aspartate-4-semialdehyde:NADP+ oxidoreductase (phosphorylating)" (EC 1.2.1.11)

R02412 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "ATP:shikimate 3-phosphotransferase" (EC 2.7.1.71)

R02413 KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzymes "Shikimate:NADP+

	3-oxidoreductase" (EC 1.1.1.25) and "L-quinate:NAD(P)+ 3-oxidoreductase" (EC 1.1.1.282)
R03083	KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "3-deoxy-D-arabino-hept-2-ulosonate-7-phosphate phosphate-lyase (cyclizing; 3-dehydroquinate-forming)" (EC 4.2.3.4)
R03084	KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "3-dehydroquinate hydro-lyase (3-dehydroshikimate-forming)" (EC 4.2.1.10)
R03460	KEGG identifier for the reaction in *Escherichia coli* catalysed by the enzyme "Phosphoenolpyruvate:3-phosphoshikimate 5-O-(1-carboxyvinyl)-transferase" (EC 2.5.1.19)
Reaction subsets	Also called enzyme subsets, sets of reactions always operating together in fixed flux proportions
S6	Ribosomal protein involved in translation
SBML	Systems Biology Markup Language (see http://sbml.org)
SBW	The Systems Biology Workbench, modular open source framework for systems biology (see http://www.sys-bio.org/)
SNA	Toolbox for analysing the possible steady-state behaviour of metabolic networks (see http://www.bioinformatics.org/project/?group_id=546)
SufS	Reaction in *Escherichia coli* catalysed by the enzyme "L-cysteine:[enzyme cysteine] sulfurtransferase" (EC 2.8.1.7); the reaction was taken from the EcoCyc database, but meanwhile it is also included in the KEGG database.

Synthetic lethality	For a pair of synthetic lethal genes the knockout of both genes is fatal while a single knockout of either gene is not.
T-invariant	Minimal T-invariants are within the theory of Petri nets the equivalents to elementary flux modes (when no reversible reactions exist).
TCA cycle	Citrate cycle; TCA cycle is the abbreviation for tricarboxylic acid cycle
TFB	Topological flux balance
Token	Demonstrative way to represent fluxes in Petri nets
UDP	Uridine 5'-diphosphate
UMP	Uridine 5'-monophosphate
UTP	Uridine 5'-triphosphate
yjeK	Also called STM4333, gene in *Salmonella enterica* encoding 2,3-β-lysine aminomutase

Chapter 1

General Introduction

Probleme kann man niemals
mit derselben Denkweise lösen,
durch die sie entstanden sind.
Albert Einstein

1.1 Motivation

Systems biology is the coordinated study of biological systems by (1) investigating the components of cellular networks and their interactions, (2) applying experimental high-throughput and whole-genome techniques, and (3) integrating computational methods with experimental efforts. This systematic approach to biology is not new, but due to progressing experimental methods and increasing computational capacities, it has attained new attraction (Kitano, 2002a; Klipp *et al.*, 2009). Historically, systems analysis was conducted in many areas of biology, such as ecology, developmental biology, and immunology. Recently, the enormous progress in genomics has moved the field of molecular biology into the area of systems biology (Westerhoff and Palsson, 2004). Present-day systems biology has two origins (Westerhoff and Palsson, 2004). On one side, there is the more familiar "biological root" comprising, for instance, the discovery of DNA and its structure in the 1950s, the development of recombinant and sequencing technologies in the 1960s–1980s, up to high-throughput technologies and the successful

sequencing of the human genome in the last 10–15 years. On the other hand, there is the less known "systemical root" of systems biology, starting with non-equilibrium thermodynamics in the 1930s, the discovery of feedback regulations in metabolism in the 1950s, the development of *metabolic control analysis* (MCA) in the 1970s (Kacser and Burns, 1973; Heinrich *et al.*, 1977), the first *in silico* models of red blood cells in the 1970s and 1980s (Rapoport *et al.*, 1974, 1976; Werner and Heinrich, 1985), up to the contemporary genome-scale models (also called whole-cell models) (Westerhoff and Palsson, 2004).

Together with the rise of systems biology also the holistic view on complex systems has been rediscovered (Kitano, 2002a,b; Cornish-Bowden *et al.*, 2004; Cornish-Bowden and Cárdenas, 2005; Cornish-Bowden, 2006). Since complex interactions of components in a biological network can lead to counterintuitive behaviour of the network, it is no longer sufficient to analyse components and interactions separately. Nevertheless, also the study of certain motifs in biological networks is important (Alon, 2007) since these occur again and again in such networks, like electronic components in a complex device, giving rise to many different overall behaviours by the various ways they are combined with each other.

Especially the massive progress in genomics, but also the considerable increase in proteomics, transcriptomics, and other "omics" data have pushed the field of molecular biology (Di Ventura *et al.*, 2006), and thus, in systems biology the analysis of biological networks, such as metabolic networks, signal-transduction networks, or gene-regulatory networks, and their integration into other cellular processes is of central importance (Schuster *et al.*, 2000a; Price *et al.*, 2004; Feist and Palsson, 2008; Salazar and Höfer, 2009).

The "holy grail" in this scientific field are kinetic models of whole cells (Tomita, 2001; Price *et al.*, 2004; Feist and Palsson, 2008) or even whole organs, such as the heart (Noble, 2002, 2006). This trend is stimulated by the excellent achievements in sequencing whole genomes. As a consequence, new software tools were developed specially designed for large systems. Examples are `Electronic cell` (Takahashi *et al.*, 2003), `Virtual Cell` (Slepchenko *et al.*, 2003), or `Silicon Cell` (Snoep *et al.*, 2006). However, most systems simulated so far by tools for whole-cell modelling have a size that can also

be handled by usual simulation packages. There is a dispute in the literature about the benefits and drawbacks of whole-cell modelling (cf. Schuster and Fell, 2007). Any model is a simplified representation of a certain aspect of reality and usually serves to answer a distinct question, e.g. what mechanism enables calcium oscillations in a cell. If the question is just about basic mechanisms and not comprising the spatial distribution of calcium ions in the cell, a two-dimensional model based on differential equations is sufficient to answer this specific issue. In contrast, the philosophy behind whole-cell modelling is different. Instead of models adapted to specific questions, a comprehensive image of the entire cell is focussed on that is at best able to answers all possible queries. Such a model would integrate all available knowledge of the structure and the parameters of the system (e.g. kinetic parameters of enzymes). It is questionable whether such a comprehensive, perfect picture can be established. Since the questions that shall be asked to a whole-cell model are previously unknown, one could even argue that the only "whole-cell model" that is sufficiently precise to answer all possible questions, is the real cell itself. But even for imperfect whole-cell models three major problems arise: (i) For the vast majority of enzymes, the kinetic parameters such as maximal velocities are unknown. In such cases, approximation methods that are only feasible if the percentage of unknown parameters is not too large, need to be used. (ii) Even though modern computer technology progresses rapidly and allows already for solving thousands of differential equations and dealing with large problems of linear programming, there are problems that are computationally hard (e.g. the combinatorial explosion, in the computation of *elementary flux modes* (EFMs, see Subsection 1.2.1 for details). (iii) As well, sometimes it is objected to whole-cell models that even if all data were known and the associated computer simulations were feasible, the vast extent of output data were too hard to analyse. Therefore, it is unlikely that simulations of the entire metabolism of complex cells will be achieved in the near future (Schuster and Fell, 2007).

Kinetic models give rise to qualitative as well as quantitative simulation results. The modelling techniques behind such kinetic models depend on the problem the models are applied to. Whereas deterministic *ordinary differential equations* (ODEs) (or *partial differential equations* (PDEs) in the case of

spatial modelling (Lemerle *et al.*, 2005)) are useful under the assumption of large particle numbers (as it is the case in metabolic networks), sparse particle numbers need the application of stochastic methods (Di Ventura *et al.*, 2006). Nevertheless, kinetic models have the disadvantage that precise kinetics of reactions are often partially or totally unknown (Di Ventura *et al.*, 2006). Despite the work to identify the components of the network from the sequenced genome of the organism under study, here also kinetic constants need to be determined. Since these constants depend on many parameters of the cellular environment, the reactions are operating in, they are difficult to ascertain. In the case that not too many parameters are unknown and the known ones are sufficiently precise, the missing parameters can be estimated by fitting the model to given experimental results (Kitano, 2002b; Di Ventura *et al.*, 2006; Wilkinson, 2009). However, for large kinetic models this approach is no longer feasible, since too many parameters are unidentified.

Therefore, methods for structural analyses are particularly successful (Reddy *et al.*, 1993; Schuster *et al.*, 2000a, 2002; Schilling *et al.*, 2000; Koch *et al.*, 2005a; Klamt *et al.*, 2006). These methods abstract from kinetic laws of the underlying reactions as well as from the precise inner structure of metabolites and focus on the stoichiometric structure of a metabolic network. The advantage of these approaches is that the stoichiometric structure of a biological network can be more easily identified than the corresponding kinetics. Moreover, the use of methods for structural modelling are useful anyway since it is necessary to get an overview of the network under study before the next level of complexity (in line with kinetic analyses) is introduced (Di Ventura *et al.*, 2006).

In the further introduction, I will review the most prominent methods of structural analysis and the field of robustness of biological systems. Since I exclusively use methods of structural modelling, I will not review the very complex and ramified field of kinetic modelling within this introduction. Apart from that, even a non-exhaustive review of kinetic modelling would be very extensive. Here I just list a few reviews for a first introduction: Andrews and Bray (2004); Lemerle *et al.* (2005); van Riel (2006); Di Ventura *et al.* (2006); Wilkinson (2009). Also there are too many software packages for kinetic modelling to list them completely. Beside the abovementioned tools, I

just want to mention a few more prominent programs such as `COPASI` (Hoops *et al.*, 2006) or its predecessor `GEPASI` (Mendes, 1993) that are described in more detail in Section 1.2 about structural modelling. Moreover, there is, for instance, `Berkeley Madonna™`, a very powerful software package for solving differential equations. `Berkeley Madonna™` is developed at the University of California at Berkeley (http://www.berkeley.edu/) and can be downloaded and purchased, respectively, via http://www.berkeleymadonna.com/.

1.2 Structural modelling

1.2.1 The concept of Elementary Flux Modes

A central concept in the structural analysis of metabolic networks is represented by the framework of *elementary flux modes* (EFMs, Fig. 1.1). EFMs are minimal sets of reactions that provide feasible fluxes at steady state through a metabolic network (Schuster and Hilgetag, 1994; Schuster *et al.*, 1999, 2000a, 2002). The *steady-state condition* (Eq. (1.2)) requires that the concentrations of metabolites within the considered metabolic network are balanced and thus remain constant. These metabolites are called internal metabolites. In contrast, external metabolites are "balanced by definition", meaning that they are either assumed to be buffered by a large number of reactions outside of the considered network (e.g. ATP or NADH+H^+) or to be available in an inexhaustible amount (e.g. growth media). EFMs are minimal in the sense that any further decomposition cuts all steady-state fluxes through the remaining reactions. The feasibility of EFMs refers to the thermodynamical constraints that need to be obeyed. The directions of the resulting fluxes may not be contradictory to the thermodynamically feasible directions of irreversible reactions. The calculation of EFMs is based on the *stoichiometric matrix* $\mathbf{N}$ of the considered metabolic network and on the *right null-space matrix* $\mathbf{K}^{(\mathbf{r})}$ of $\mathbf{N}$, also called *kernel* matrix.

Compared with the classical biochemical approach of representing a metabolic network as a set of metabolites and reactions, the description of a metabolic network as a stoichiometric matrix is a more abstract but also more compact form. In a stoichiometric matrix $\mathbf{N}$ the consumption and

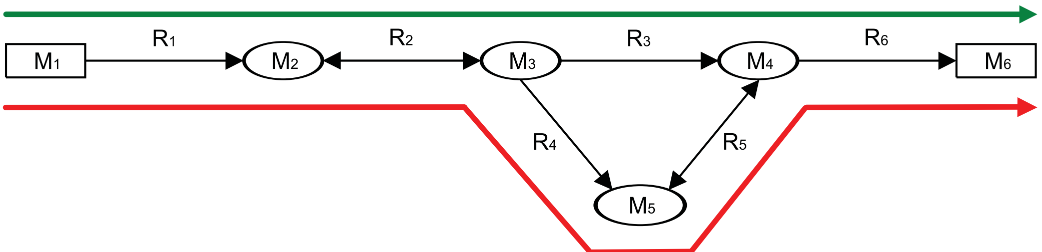

Figure 1.1. Small prototypic example of a metabolic network together with the two irreversible EFMs the system gives rise to. This example contains only monomolecular reactions. Thus, it is sufficient to represent them with usual arcs. But for the representation of bimolecular or even higher-molecular reactions the use of hyperarcs becomes necessary. External metabolites are boxed whereas internal metabolites are depicted as ellipses. The first EFM (green arrow) consists of reactions R_1–R_3 and R_6, the second (in red) of R_1–R_2 and R_4–R_6.

production of metabolites in a reaction network are summarised by rows corresponding to metabolites and columns corresponding to reactions. An entry $\mathbf{N}_{i,j}$, for instance, specifies the stoichiometric coefficient with which metabolite i is consumed or produced in reaction j. If metabolite i is a substrate of reaction j, $\mathbf{N}_{i,j}$ is negative. If i is a product of j, $\mathbf{N}_{i,j}$ is positive and in the case that metabolite i is not participating in reaction j, $\mathbf{N}_{i,j}$ is zero. Eq. (1.1) shows the stoichiometric matrix of the prototypic example depicted in Fig. 1.1. The matrix contains $n = 6$ columns corresponding to the reactions R_1–R_6 and $m = 4$ rows representing the internal metabolites R_2–R_5.

$$\mathbf{N} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 \end{pmatrix} \qquad (1.1)$$

A stoichiometric matrix alone is not a sufficient representation of a metabolic network because of several reasons. First, as can be seen from Eq. (1.1), the external metabolites that are balanced by definition, are not part of the stoichiometric matrix. Second, the stoichiometric matrix does not contain information about the reversibility or irreversibility of reactions. And third, stoichiometric coefficients of components participating as catalysts in reactions and thus being present on the educt and product sides of theses reactions, are cancelled out to zero. Hence, in the case of representing a

metabolic system by a stoichiometric matrix, these missing data must be stored separately to have a full complement for the representation of the network by a set of reaction equations.

The right null-space matrix $\mathbf{K^{(r)}}$ of $\mathbf{N}$ is defined by the equation

$$\mathbf{NK^{(r)}} = \mathbf{0}, \tag{1.2}$$

with $\mathbf{0}$ being a zero matrix of appropriate dimension. Eq. (1.2) is the mathematical definition of the steady-state condition. As already mentioned, it implies that the consumption and production of internal metabolites must be balanced. An important characteristic of the kernel matrix $\mathbf{K^{(r)}}$ is that its columns are vectors spanning the so-called *right null space* with all valid flux distributions lying inside or at the edges of this space. The dimension of the kernel matrix $\mathbf{K^{(r)}}$ is $n \times d$, where n is the number of reactions (columns in $\mathbf{N}$) and d the dimension of the right null space (equal to the number of reactions minus the rank of $\mathbf{N}$):

$$d = n - rank(\mathbf{N}) \tag{1.3}$$

The kernel matrix for the prototypic example presented in Fig. 1.1 reads as follows:

$$\mathbf{K^{(r)}} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \tag{1.4}$$

It can be easily seen that in this case the columns of the kernel matrix shown in Eq. (1.4) correspond already to the two EFMs that are depicted in Fig. 1.1 because the first column represents a flux through reactions R_1–R_3 and R_6 (green EFM in Fig. 1.1) and the second column a flux through reactions R_1–R_2 and R_4–R_6 (red EFM in Fig. 1.1). For complex systems this simple equivalence is usually not the case.

All flux distributions satisfying Eq. (1.2) are linear combinations of col-

umns of $\mathbf{K}^{(\mathbf{r})}$. Hence, the analysis of $\mathbf{K}^{(\mathbf{r})}$ can reveal meaningful relationships about fluxes at steady state. One such relationship is the concept of *reaction subsets*, also called *enzyme subsets*, that was introduced by Pfeiffer *et al.* (1999). Reaction subsets are sets of reactions that are always operating together in fixed flux proportions. They can be identified from rows in $\mathbf{K}^{(\mathbf{r})}$ that are multiples of each other. This concept was further generalised by Poolman *et al.* (2007). They are calculating so-called *reaction correlation coefficients*, a concept that is a logical extension of the concept of enzyme (or reaction) subsets. A reaction correlation coefficient ϕ_{ij} is given as the cosine of the angle $\theta_{ij}^{\mathbf{K}^{(\mathbf{r})}}$ between two row vectors K_i and K_j of $\mathbf{K}^{(\mathbf{r})}$:

$$\phi_{ij} = \cos\left(\theta_{ij}^{\mathbf{K}^{(\mathbf{r})}}\right) \tag{1.5}$$

As a result, ϕ_{ij} is one if $\theta_{ij}^{\mathbf{K}^{(\mathbf{r})}} = 0$, meaning that K_i and K_j are parallel and thus are multiples of each other (which corresponds to the classical definition of a reaction subset given by Pfeiffer *et al.* (1999). In contrast, ϕ_{ij} is zero if $\theta_{ij}^{\mathbf{K}^{(\mathbf{r})}} = \pi/2$, which implies that K_i and K_j are orthogonal and thus completely independent from each other. Values of ϕ_{ij} between one and zero correspond to partially coupled fluxes.

The analysis of the flux vectors contained in $\mathbf{K}^{(\mathbf{r})}$ is restricted by two reasons. On one hand, $\mathbf{K}^{(\mathbf{r})}$ is not unique, meaning that there is an infinite number of matrices representing a right null space of $\mathbf{N}$. On the other hand, the analysis of $\mathbf{K}^{(\mathbf{r})}$ is based on the assumption that all reactions are reversible. But due to thermodynamic reasons, some reactions might be practically irreversible at physiological conditions. We therefore need to introduce a subvector V_{irr} of flux vector V that comprises only the fluxes through the irreversible reactions. V_{irr} is defined by decomposition of V in the following way:

$$V = \begin{pmatrix} V_{rev} \\ V_{irr} \end{pmatrix} \tag{1.6}$$

Now, V_{irr} has to fulfil the following *irreversibility condition* (also called *sign restriction*) in order to let all irreversible reactions be thermodynamically possible:

$$V_{irr} \geq \mathbf{0}, \tag{1.7}$$

meaning that all elements v_k of V_{irr} need to be greater or equal zero. The steady-state condition and the irreversibility condition (Eqs. (1.2) and (1.7)) form a linear equation/inequality system. The solution of this system corresponds to a convex polyhedral cone in the n-dimensional flux space (Fig. 1.2) with n being the number of columns of the kernel matrix. The edges of this cone correspond to EFMs. Additional EFMs may lie inside this cone.

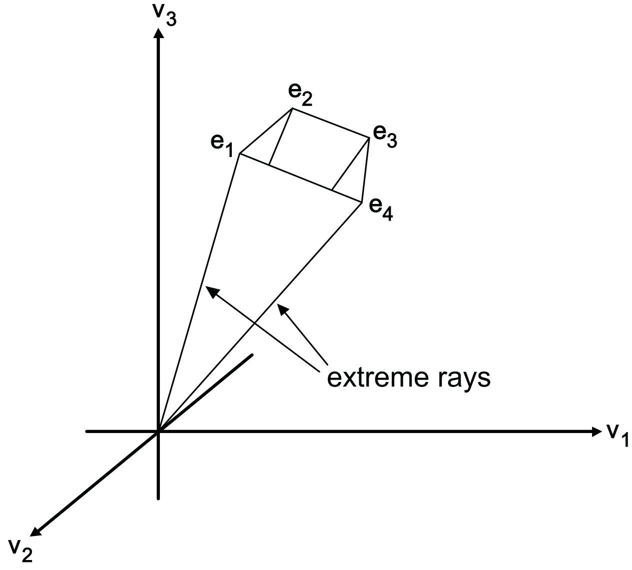

Figure 1.2. One possible solution space of the steady-state condition and the irreversibility condition (Eqs. (1.2) and (1.7)) in a 3-dimensional flux space. The presented solution space is a convex polyhedral cone that is spanned by the four so-called extreme rays e_1–e_4 which are EFMs of the corresponding system.

For the exact definition of an *elementary flux mode* (EFM), first a *flux mode* needs to be defined (cf. Schuster *et al.*, 2002).

Definition 1.1 (flux mode)
A flux mode, M, in an n-dimensional flux space is defined as the set

$$M = \{V \in \mathbb{R}^n : V = \lambda V^*, \lambda > 0\}, \tag{1.8}$$

where V^* is an n-dimensional vector (unequal to the null vector) fulfilling the following two conditions:
(C1) *Steady-state condition, i.e. Eq. (1.2).*
(C2) *Irreversibility condition, i.e. V^* contains a subvector V^*_{irr} that fulfils*

inequality Eq. (1.7).

Now, elementary flux modes can be defined.

Definition 1.2 (elementary flux mode)
For any flux vector, V, with elements v_k, let the support of V be the set of positions where V is not zero:

$$supp(V) = \{i : v_i \neq 0\}. \tag{1.9}$$

A flux mode M^* is called an elementary flux mode *if, and only if, M^* fulfils the following* non-decomposability condition *requiring that there exists no other flux mode M^{**} (unequal to the null vector), the support of which would be a proper subset of the support of M^*:*

$$\nexists\, M^{**} : supp(M^{**}) \subset supp(M^*) \tag{1.10}$$

Simply speaking, the definition of EFMs means that no EFM, M^{**}, can be derived from another EFM, M^*, by removing one or more reactions. As a consequence of the non-decomposability condition (also referred to as *elementarity condition*) the elementarity of all flux modes, found inside the solution space, needs to be tested. If they fulfil the non-decomposability condition, these flux modes are (additional) EFMs. As already mentioned, for the small example shown in Fig. 1.1, the columns of the kernel matrix (Eq. (1.4)) are already the two depicted EFMs. But for larger systems the check for elementarity is a computational demanding task.

For this purpose, sophisticated test methods are necessary. The most common algorithms for the computation of EFMs are the convex base approach, introduced by Schuster *et al.* (2000a) and the null-space approach, developed by Urbanczik and Wagner (2005). Since the metabolic networks under study are continuously increasing in size, the algorithms for the computation of EFMs need to be further improved. The latest and currently fastest algorithm (an improved version of the original null-space implementation), presented by Terzer and Stelling (2008), is currently able to enumerate more than 26 million EFMs within about three hours (without pre- and postprocessing steps) on a Linux machine with two dual-core AMD Opteron

processors and 30GB of memory (see supplementary material of Terzer and Stelling (2008)). This enormous number points to the limitations of EFM analysis. As Klamt and Stelling (2002) demonstrated, the number of EFMs might grow exponentially with the network size. This restricts the applicability of EFM analyses to medium-sized metabolic models of around 150 reactions and metabolites. Although there are already methods under development to determine at least subsets of the EFMs in genome-scale metabolic networks (networks covering the metabolism of an entire organisms), such as *elementary flux patterns* (Kaleta *et al.*, 2009) or *K-shortest elementary flux modes* (de Figueiredo *et al.*, 2009a), the enumeration of all EFMs in a genome-scale network is still not possible. Another challenge in EFM analysis is the further processing (filtering, clustering etc.) of the huge amount of resulting EFMs. This step requires sophisticated and automated methods to make comprehensive EFM analyses in large metabolic networks practicable.

Despite the limitations of EFM analysis in genome-scale networks this concept has a wide field of application. The scope ranges from pathway prediction (Liao *et al.*, 1996; Schuster *et al.*, 1999; Fischer and Sauer, 2003), use in functional genomics (Pachkov *et al.*, 2007), assessment of network flexibility (Stelling *et al.*, 2002) up to biotechnological applications (Carlson and Doyle, 2002; Schwender *et al.*, 2004). A pretty useful application that is based on EFMs is the concept of *minimal cut sets* (MCSs, Klamt and Gilles, 2004; Klamt, 2006) which allows for determining smallest (irreducible) sets of interventions (e.g. knockouts of reactions or missing metabolites) to block certain targets (so-called *objective functions* such as all EFMs in a system that are producing amino acids. For a review on the various applications of EFMs see Trinh *et al.* (2009).

For computing EFMs we use the `Metatool` software, developed earlier in our group (Pfeiffer *et al.*, 1999; von Kamp and Schuster, 2006). But there is a lot of other software packages for the computation and analysis of EFMs. Here I give just a short and incomplete list of the most prominent tools.

- `Metatool` (Pfeiffer *et al.*, 1999; von Kamp and Schuster, 2006) is available in two versions. Since version 5.0 `Metatool` is designed as `MATLAB` software package and implements the null-space approach, developed by Urbanczik and Wagner (2005) but with an algebraic test of candidate

modes for their mutual independence. Older versions (up to 4.9) are compiled as stand-alone executables. `Metatool` imports metabolic networks represented in the so-called *Metatool format* (for details and for the download of `Metatool` see `http://pinguin.biologie.uni-jena.de/bioinformatik/networks/index.html`). The output can be accessed via text file or via further analysis of the data that `Metatool` returns to the `MATLAB` workspace. If the SBML Toolbox for `MATLAB` is installed (Keating *et al.*, 2006) `Metatool` is able to import also metabolic networks in the SBML format (Hucka *et al.*, 2003).

- `efmtool` (Terzer and Stelling, 2008) is a Java-based library that offers also a `MATLAB` integration. `efmtool` processes metabolic networks in the SBML format. It implements the currently fastest algorithm for the computation of EFMs which is an improved version of the null-space approach and parallelised in that way that it is possible to exploit multi-core processor architectures. `efmtool` can be downloaded from `http://www.csb.ethz.ch/tools/efmtool/`.

- `CellNetAnalyzer` (Klamt *et al.*, 2006, 2007) (the successor of FluxAnalyzer (Klamt *et al.*, 2003)) is a `MATLAB` software package for EFM analysis with a graphical user interface. The user interface provides the possibility to display EFMs within ready-made representations of the metabolic networks. For computing EFMs, `efmtool` or `Metatool` serve as back-ends. Many additional tools such as the computation of MCSs and the analysis of Boolean networks are included. `CellNetAnalyzer` can be downloaded from `http://www.mpi-magdeburg.mpg.de/projects/cna/cna.html`.

- `YANAsquare` (Schwarz *et al.*, 2007) is a Java-based program being developed by the group of Prof. Dandekar at the University of Würzburg. It comprises a graphical user interface and integrates tools for network reconstruction such as import of reactions and metabolites from the KEGG database (Kanehisa *et al.*, 2008), automated network visualization and EFM analysis. Reaction networks can be opened and saved either in `Metatool` format or in SBML format. The EFMs are

computed by a built-in stand-alone version of `Metatool` or a similar Java-based implementation. `YANAsquare` can be downloaded from `http://yana.bioapps.biozentrum.uni-wuerzburg.de/`.

- `COPASI` (Hoops *et al.*, 2006) (the successor of `GEPASI` (Mendes, 1993)) is a software package for simulation and analysis of biochemical networks and their dynamics. It is designed as stand-alone program that supports network models in the SBML format. Its main focus is on dynamic simulations. But besides features such as simulating network behaviour using ODEs, stochastic simulations, or stability analysis, `COPASI` is also able to calculate EFMs. `COPASI` can be downloaded from `http://www.copasi.org`.

- `SBW`, *The Systems Biology Workbench* (Bergmann and Sauro, 2006) is a modular open source framework for systems biology connecting heterogeneous software applications that is developed by the group of Herbert Sauro at the University of Washington. The software framework supports SBML format for data exchange between the different modules. Some prominent modules are `Jarnac` (successor of `SCAMP`), a fast simulator of reaction networks, `JDesigner`, a user-friendly GUI front end to an SBW compatible simulator, and an SBW-compatible version of `Metatool`. `SBW` can be downloaded from `http://www.sys-bio.org/`.

- `ScrumPy` (Poolman, 2006) is a software package for definition and analysis of metabolic models and was developed in the Cell Systems Modelling Group of Prof. Fell at the Oxford Brookes University. It is written in Python and this programming language is also implemented as user interface. Network models are managed as SBML files. The program offers features for both kinetic and structural modelling, but with the emphasis on the latter one. One important feature is the computation of so-called *elementary leakage modes* in order to detect stoichiometric inconsistencies in large metabolic systems (Gevorgyan *et al.*, 2008). `ScrumPy` can be downloaded from `http://mudshark.brookes.ac.uk/index.php/Software/ScrumPy`.

- `PySCeS`, *The Python Simulator for Cellular Systems* (Olivier *et al.*,

2005) is a software package providing a variety of tools for the analysis of cellular systems, such as functionalities for structural network analysis, kinetic simulation of a cellular systems, Metabolic Control Analysis, bifurcation analysis, and parameter scans. The software package is developed by the Triple-J Group for Molecular Cell Physiology at the Stellenbosch University in South Africa (`http://www.jjj.sun.ac.za/`). It is written for Python and network models can be managed as SBML files. `PySCeS` can be downloaded from `http://pysces.sourceforge.net/`.

- `SNA` (Urbanczik, 2006) is an interactive toolbox for analysing the possible steady-state behaviour of metabolic networks by computation and enumeration of the generating and elementary vectors of their flux and conversions cones. It also supports analysing the steady states by linear programming (flux balance analysis). The toolbox is implemented mainly in `Mathematica` and returns numerically exact results. `SNA` is released under an open source license and can be downloaded from `http://www.bioinformatics.org/project/?group_id=546`.

Not only the analysis of the right null-space matrix $\mathbf{K^{(r)}}$ of $\mathbf{N}$ (Eq. (1.2)) can reveal meaningful results. Beside the kernel matrix there is also the *left null-space matrix* $\mathbf{K^{(l)}}$ of $\mathbf{N}$ that is defined by the equation

$$\mathbf{K^{(l)}N} = \mathbf{0}. \tag{1.11}$$

The rows of the left null-space matrix are vectors spanning the so-called *left null space* that contains all linear combinations of metabolite concentrations of a system that always remain constant. In contrast to the steady-state condition that implies that the consumption and production of internal metabolites must be balanced, this equation describes the basic principle of mass conservation within closed systems, which requires that the sum of atoms, forming the metabolites, must be the same on both sides of a reaction equation. The columns of $\mathbf{K^{(l)}}$ correspond to particular metabolites whereas the rows represent linear combinations of so-called *mass conservation relations* that are also referred to as *conserved moieties* (Schuster and Höfer, 1991).

If a column of $\mathbf{K^{(1)}}$ contains only zeros, the corresponding metabolite does not participate in any conservation relation of the considered system. This means that some reactions of the metabolic network do not fulfil the mass conservation and thus there exist reactions consuming or producing mass. Obviously, at this point external metabolites need to be included in the stoichiometric matrix $\mathbf{N}$. Otherwise, a reaction equation like $A \rightarrow B$ with A being an external metabolite would result in a reaction that produces B from nothing, which apparently violates the principle of mass conservation.

The full stoichiometric matrix $\mathbf{N}^*$ of the example given in Fig. 1.1 includes the external metabolites M_1 (first row) and M_6 (last row):

$$\mathbf{N}^* = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.12}$$

The left null-space matrix $\mathbf{K^{(1)}}^*$ of $\mathbf{N}^*$ reads as follows:

$$\mathbf{K^{(1)}}^* = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{1.13}$$

Analogously to the kernel matrix, the dimension of the left null-space matrix $\mathbf{K^{(1)}}^*$ is $m \times d$ with m being the number of metabolites (rows in $\mathbf{N}^*$) and d being now the dimension of the left null space (equal to the number of metabolites minus the rank of $\mathbf{N}^*$):

$$d = m - rank(\mathbf{N}^*) \tag{1.14}$$

The matrix $\mathbf{K^{(1)}}^*$ can be easily translated into the following conservation relation that illustrates clearly the conservation of mass in the considered system.

$$[M_1] + [M_2] + [M_3] + [M_4] + [M_5] + [M_6] = const. \tag{1.15}$$

with $[M_1]$–$[M_6]$ being, for instance, the concentrations or the amounts of substance of the metabolites M_1–M_6. In more complex systems $\mathbf{K^{(l)}}$ may also contain negative entries which are difficult to interpret in the sense of atomic compositions. But just as the right null-space matrix $\mathbf{K^{(r)}}$ also the left null-space matrix $\mathbf{K^{(l)}}$ is not unique. As a consequence, similar to EFMs, new conservation relations can be derived as linear combinations of the existing ones. It was shown by Schuster and Höfer (1991) that within closed systems (systems without external metabolites), from each set of conservation relations another set with the same number of conservation relations can be deduced, the vectors of which no longer comprise negative entries.

One application of this concept of mass conservation is the detection of stoichiometric inconsistencies in metabolic networks. For instance, the left null-space matrix $\mathbf{K^{(l)}}$ of the stoichiometric matrix $\mathbf{N}$ is empty. This means that the system shown in Fig. 1.1 does not contain any conserved moiety when the metabolites M_1 and M_6 are defined as external since the reactions R_1 and R_6 are no longer mass-balanced. But of course, apart from these "stoichiometric inconsistencies" given by the defintion of external metabolites, substantial stoichiometric errors can be found by the analysis of the left null-space matrix. This application is feasible even in large networks since, in contrast to EFMs, conservation relations can also be computed in genome-scale metabolic networks (Schuster and Höfer, 1991). Gevorgyan *et al.* (2008) computed so-called *elementary leakage modes* in order to detect reactions that do not obey the principle of mass conservation. Such a mode indicates a set of reactions by which mass can be produced or is consumed. Sometimes stoichiometric inconsistencies are induced by simplifications in the model, such as neglecting small metabolites (e.g. O_2, CO_2, or N_2).

1.2.2 Petri Nets

The concept of *Petri nets* was introduced in 1962 by the German mathematician and computer scientist Carl Adam Petri (Petri, 1962, cf. Starke (1990)). Although Petri nets were originally devised for the description of production and transportation processes they are also applied to biochemical reaction systems. The first approaches in that context were started by Reddy *et al.*

(1993) and Hofestädt (1994). Beside the huge number of Petri-net applications to modelling metabolic networks (Zevedei-Oancea and Schuster, 2003; Koch *et al.*, 2005a,b) also an increasing usage in modelling signalling networks and gene-regulatory networks can be found in the literature (Matsuno *et al.*, 2000; Sackmann *et al.*, 2006; Kielbassa *et al.*, 2009).

Petri nets are bipartite graphs consisting of two types of nodes, places and transitions, connected by arcs (directed edges). In metabolic networks, these nodes represent metabolites and reactions, respectively. Transitions in Petri nets are unidirectional. Therefore reversible reactions need to be decomposed into forward and backward reaction steps. A notable property of Petri nets (in comparison to other bipartite graphs) is the concept of tokens. Tokens (usually represented by solid dots) are discrete representations of those objects that move along transitions from one place to the next one. This token flow along an active transition is often called "firing". In the context of metabolic reaction networks tokens usually represent molecules. The distribution of tokens over all places at a given time point is called marking. The arcs can be weighted (in graphical representations indicated by a number next to the arc) to represent, for instance, stoichiometries of reactions (usually only if they are greater than one). The matrix that is comprising all these coefficients is called *incidence matrix*, $\mathbf{B}$, and is equivalent to the stoichiometric matrix $\mathbf{N}$ of a metabolic network.

An advantage of Petri nets is the better visualisation of fluxes by token flows and the easy usage of sophisticated software tools (Sackmann *et al.*, 2006). A drawback is the more laborious representation of hyperarcs by two types of nodes. On the other hand there are many graph-drawing tools that still cannot handle hyperarcs making the bipartite representation of hyperarcs necessary again. Tokens are a demonstrative way to represent fluxes but their discreteness is usually unnecessary since (at least in metabolic networks) the numbers of molecules are very high. Consequently, Petri nets might be more applicable for systems such as gene-regulatory and signalling networks that typically comprise smaller numbers of molecules. Nevertheless, there are extensions to continuous versions of Petri nets, so-called *hybrid functional Petri nets* (Matsuno *et al.*, 2000; Wu and Voit, 2009).

The interest in structural properties of Petri nets emerged indepen-

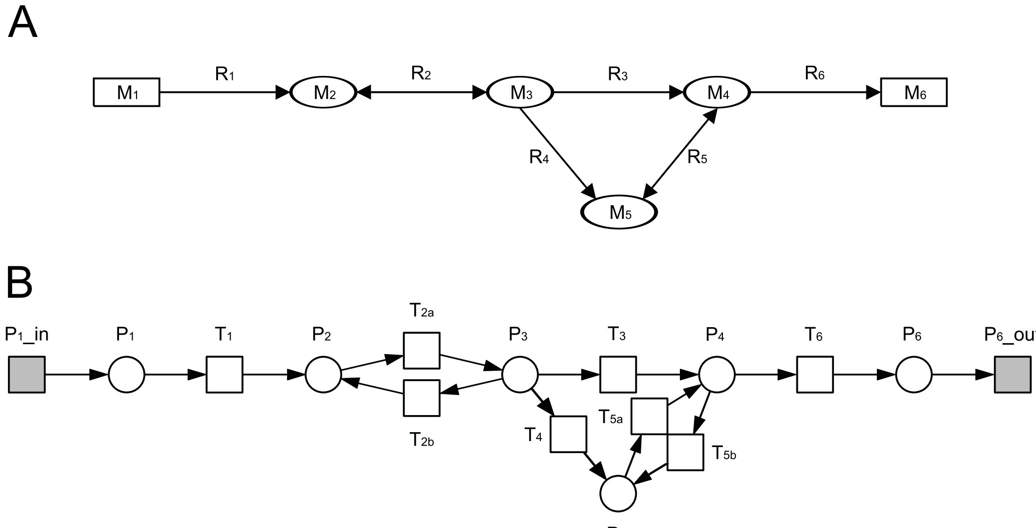

Figure 1.3. Different representations of the small prototypic example shown in Fig. 1.1. **A** Usual representation with arcs/hyperarcs. External metabolites are boxed whereas internal metabolites are depicted as ellipses. **B** Representation of the same system as a Petri net. Here the transitions T_1–T_6 are displayed as squares and the places P_1–P_6 as circles. The places P_1 and P_6 are source and sink, meaning that they are generating and consuming infinite numbers of tokens, respectively. To account for theses properties, two boundary transitions P_1_in and P_6_out are introduced. The generating pre-transition P_1_in produces P_1 from nothing and the consuming post-transition P_6_out converts P_6 to nothing. Reversible reactions are split up into forward and backward reactions. The T-invariants the Petri net gives rise to, correspond to the two EFMs that are depicted in Fig. 1.1.

dently of the structural analysis of metabolic networks. In Petri net theory, *T-invariants* and *P-invariants* have been defined by Lautenbach (1973, cf. Starke (1990)) that correspond to the concepts of EFMs and conservation relations. A vector V is called a T-invariant if it fulfils the steady-state condition (Eq. 1.2). This means that T-invariants are sets of token flows that preserve the initial marking of the Petri net. Analogous to the sign restriction for irreversible fluxes in the concept of EFMs (Eq. 1.7), T-invariants must accommodate the fact that all transitions are unidirectional. Thus, a subset of all T-invariants, the so-called true T-invariants, is defined. True T-invariants V satisfy the definition of transitions by having all components in V being non-negative and at least one component being positive. Within the subset of true T-invariants a further subset of minimal T-invariants is defined. For this definition again the mathematical concept of support is

used. A true T-invariant V^* is called a minimal T-invariant if there is no other true T-invariant, V^{**}, the support of which would be a valid subset of the support of V^*:

$$supp(V) = \{i : V_i \neq 0\} \tag{1.16}$$

$$\nexists\, V^{**} : supp(V^{**}) \subset supp(V^*) \tag{1.17}$$

This means that no minimal T-invariant, V^{**}, can be derived from another minimal T-invariant, V^*, by removing one or more transitions. This definition corresponds to the elementarity condition for EFMs.

In the case, where no reversible reactions exist or all of them have been split into irreversible forward and backward steps, the sets of minimal T-invariants and EFMs are equivalent. Otherwise, the set of EFMs differs from the set of minimal T-invariants by lacking those futile cycles that only consist of the forward and backward steps of the corresponding reversible reactions. Thus in Petri nets these spurious (and thermodynamically impossible) cycles need to be discarded after computation.

Corresponding to the conserved moieties in biochemical networks that are derived from the left null space, in the theory of Petri nets P-invariants have been introduced (Lautenbach, 1973, cf. Schuster *et al.* (2000b)). Whereas conservation relations describe mass conservations of a certain system, P-invariants specify relations of marking conservation. That means that the sum of token numbers of a certain set of places remains constant independent of any firing transitions. The definition of minimal P-invariants is analogously to that of minimal T-invariants. A discussion of the similarities between EFMs and minimal T-invariants as well as between conserved moieties and minimal P-invariants was first published by Schuster *et al.* (2000b) and by Voss *et al.* (2003).

Another relation between EFMs and Petri nets is the similarity between the concept of enzyme subsets that was already discussed above and the concept of maximal common transition sets (MCT-sets) that was introduced by Sackmann *et al.* (2006). They group the transitions of a Petri net into so-called maximal common transition sets (MCT-sets) by their occurrence in

the minimal T-invariants. For instance, the transitions t_i and t_j are grouped into the same MCT-set if and only if they participate (together) in exactly the same minimal T-invariants. This means that an MCT-set is a maximal set of transitions that operate together in each steady state and can be interpreted as a building block of the Petri net. As it is the case for reactions within enzyme subsets, the transitions belonging to one MCT-set need not be adjacent to each other. But nevertheless, the definition of enzyme subsets is more strict because it additionally requires fixed flux proportions between participating reactions. Thus, all enzyme subsets correspond to MCT-sets but not vice versa.

For editing and analysing Petri nets, there is a lot of software packages around. Here I give just a short list of programs used in our group.

- `Snoopy` (Heiner *et al.*, 2008; Rohr *et al.*, 2010) is a software package to construct and animate hierarchical graphs, especially Petri nets. The program has been developed at the University of Technology in Cottbus. `Snoopy` can be used to validate technical systems, as well as biochemcial networks such as metabolic, signal-transduction, or gene-regulatory networks. `Snoopy` offers an intuitve graphical user interface to design Petri nets and to simulate their token flows in an animated manner. The support for hierarchical graphs (hierarchical nodes, logical nodes) is particularly useful for larger models, or models with an higher connectivity degree. `Snoopy` provides import from the several tools (e.g. PED) and file formats (e.g. SBML (Hucka *et al.*, 2003)). Its export functionality comprises a lot of analysis tools and file formats. I just list up here Charlie, INA, and again SBML. The predecessor of `Snoopy` is the hierarchical Petri-net editor `PED` `http://www-dssz.informatik.tu-cottbus.de/index.html?/software/ped.html`. `Snoopy` can be downloaded from `http://www-dssz.informatik.tu-cottbus.de/index.html?/software/snoopy.html`.

- `Charlie` `http://www-dssz.informatik.tu-cottbus.de/index.html?/software/charlie.html` is a software package to analyse Petri nets. The tool was developed at the University of Technology in

Cottbus. `Charlie` was designed according to experiences from the use of the program `Integrated Net Analyzer (INA)`. It is implemented in Java and offers a graphical user interface, a textual user interface and more features, such as a marking editor. `Charlie` is able to import all Petri-net classes that are supported by `Snoopy`. The tool is in use to check technical systems, as well as biological networks. Among its features are functions for the analysis of structural properties such as deadlocks/traps or P- and T-invariants, determining shortest paths, and the visualisation of reachability/coverability graphs.

- `Integrated Net Analyzer (INA)` http://www2.informatik.hu-berlin.de/lehrstuehle/automaten/ina/ is a tool package for the analysis of Petri nets and Coloured Petri nets that was developed in the group of Prof. Peter H. Starke at the Humboldt University in Berlin. `INA` comprises a textual editor to compile and to edit nets, a simulation function, and an analysis function to compute structural information, P- and T-invariants, and reachability/coverability-graphs. The simulation function allows for starting at a given marking and to forward tokens along single transitions or maximal steps; the user can thus traverse parts of the reachability graph. The analysis can be carried out under different transition rules, with or without priorities or time restrictions, and under firing of single transitions or maximal sets of concurrently enabled transitions. The analysis of invariants allows for the computation of generator sets of all P- and T-invariants and of all non negative invariants. Furthermore, minimal pathways can be computed, and the (non)reachability of a certain marking can be determined. Some external graphical editors and tools can export nets to INA, such as `Snoopy` or its predecessor `PED`.

For a comprehensive list on available Petri-net tools see http://www.informatik.uni-hamburg.de/TGI/PetriNets/tools/db.html.

1.2.3 Chemical Organisation Theory

Chemical Organisation Theory (Dittrich and Speroni di Fenizio, 2007) has recently been introduced as another tool to analyse capacities and structure of metabolic networks. This concept offers the possibility to decompose a metabolic network into (algebraically) closed and self-maintaining subnetworks that form a hierarchy. The hierarchy of organisations can be visualised by displaying them in an Hasse diagram (Kaleta *et al.*, 2006). One important conclusion that can be deduced from organisation theory is that every steady state of a metabolic network can be mapped to an organisation (Dittrich and Speroni di Fenizio, 2007). Although this concept is not applied within this thesis, I want to summarise it shortly, on one hand for the sake of completeness and on the other hand to point out its relations to the concept of EFMs.

The central terms of Chemical Organisation Theory can be defined by the following four definitions (cf. Kaleta *et al.*, 2006).

Definition 1.3 (Algebraic Chemistry)
Let $\mathcal{M}$ be a set of elements (called species, molecular species, or just molecules). Then $\mathcal{P}_{\mathcal{M}}(\mathcal{M})$ denotes the set of all multisets *with elements from $\mathcal{M}$. A multiset differs from a common set in the fact that it can comprise the same element more than once. Reactions occuring among species $\mathcal{M}$ can then be defined by a relation $\mathcal{R} : \mathcal{P}_{\mathcal{M}}(\mathcal{M}) \times \mathcal{P}_{\mathcal{M}}(\mathcal{M})$. The pair $\langle \mathcal{M}, \mathcal{R} \rangle$ is called an* algebraic chemistry*.*

Definition 1.4 (Closed Set)
A set of species $S \subseteq \mathcal{M}$ is closed *if for all reactions $(A \rightarrow B) \in \mathcal{R}$, with $A \in \mathcal{P}_{\mathcal{M}}(S) \Rightarrow B \in \mathcal{P}_{\mathcal{M}}(S)$.*

Simply speaking, if S comprises the educts of a reaction, then S must also contain all corresponding products. There is no reaction that can convert species in S to any new species not being in S.

Definition 1.5 (Self-Maintaining Set)
Given an algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$ with $m = |\mathcal{M}|$ species and $n = |\mathcal{R}|$ reactions, its dynamics can be described by $\dot{C} = \mathbf{N}V$ with concentrations

vector $C \in \mathbb{R}_+^m$*, stoichiometric matrix* $\mathbf{N}$*, and flux vector* $V \in \mathbb{R}_+^n$*. A set of species* $S \subseteq \mathcal{M}$ *is called* self-maintaining *if a flux vector* V *exists, so that the following conditions are fulfilled:*

1. *For every reaction* $(A \rightarrow B) \in \mathcal{R}$ *with* $A \in \mathcal{P}_\mathcal{M}(S)$*, its corresponding flux is* $V_{A\rightarrow B} > 0$*.*

2. *For every reaction* $(A \rightarrow B) \in \mathcal{R}$ *with* $A \notin \mathcal{P}_\mathcal{M}(S)$*, its corresponding flux is* $V_{A\rightarrow B} = 0$*.*

3. *For every species* $i \in S$*, its concentration change is non-negative:* $\dot{C}_i = (\mathbf{N}V)_i \geq 0$*.*

This definition means that by considering only the subnetwork spanned by the species in S and additionally those that can be produced from these, a positive flux vector can be found such that no species of S is depleted.

Definition 1.6 (Organisation)
A set of species $S \subseteq \mathcal{M}$ *that is closed and self-maintaining is called an* organisation*.*

The connection between the concept of EFMs and Chemical Organisation Theory is the following: each steady state of a system is on one hand an organisation and can be interpreted on the other hand as a combination of EFMs (Kaleta *et al.*, 2006). In contrast to EFMs, organisations are defined as sets of species, not reactions. The property of self-maintenance points out a relation between EFMs and chemical organisations: Whereas the union of all solution spaces of all species sets lies within a convex cone in flux space that is defined by the self-maintenance condition $\mathbf{N}V \geq 0$ related to all m species in $\mathcal{M}$ and the restriction to non-negative fluxes $V \geq 0$ (cf. definitions above), the equalities/inequalities that represent all steady-state flux distributions of the system are more strictly defined as $\mathbf{N}V = 0$ and $V_{irr} \geq 0$. Hence, obviously all flux vectors (apart from all zero-component vectors) fulfilling the steady-state condition also fulfil the self-maintenance condition. Consequently, the steady-state cone lies within the self-maintenance cone except for null vectors that are only allowed by the steady-state condition. This leads to the following five conclusions (cf. Kaleta *et al.*, 2006).

1. In a system where all metabolites decay spontaneously, all organisations can be found by using the convex cone that is spanned by the EFMs.

2. The concept of EFMs can be applied to search for organisations fulfilling the steady state condition. Such steady-state organisations are combinations of EFMs.

3. An EFM consists of a unique set of organisations. The smallest organisations containing the EFM constitute this set.

4. Organisations don't need to contain EFMs because they can also imply the accumulation of metabolites.

5. Since EFMs can start or end with external and thus unbalanced metabolites, the set of metabolites belonging to an EFM is not necessarily self-maintaining.

1.2.4 Flux Balance Analysis

Flux balance analysis (FBA) is an application of the mathematical concept of linear optimisation to metabolic networks obeying steady state. It was first described (and called FBA) by Savinell and Palsson (1992a,b) although there is an earlier study of fat synthesis in adipose tissue by Fell and Small (1986) where FBA was applied (but not yet named FBA). FBA allows for the determination of steady-state fluxes V on the basis of the stoichiometric matrix $\mathbf{N}$.

$$\mathbf{N}\mathbf{K}^{(\mathbf{r})} = 0 \tag{1.18}$$

with $\mathbf{K}^{(\mathbf{r})}$ being again the right null space containing all steady-state fluxes V in the system.

This equation system is linear and homogeneous. In the case of metabolic networks, it is usually underdetermined, meaning that there is an infinite number of flux distributions fulfilling the equation system. To constrict the solution space additional constraints are introduced. Additional constraints can be, for instance, the irreversibility of reactions or minimal and maximal flux values. To obtain discrete flux distributions linear optimisation is performed. Assuming the case that the maximum possible cell-growth yield on

a medium with limited glucose source is of interest, the objective function of the optimisation problem is the weighted sum of all biomass components that are produced from glucose. The result is the amount of biomass produced from one molecule of glucose. Ideally, the result is one unique flux distribution, but very often this is not the case. In such cases, it is necessary to calculate intervals for each flux (Mahadevan and Schilling, 2003). The objective function can be chosen arbitrarily, for instance to determine the maximum possible production rate of a certain amino acid.

The concept has been expanded by several variations, such as *minimization of metabolic adjustment* (Segrè *et al.*, 2002) or *energy balance analysis* (Beard *et al.*, 2002). Due to the vast number of publications concerning this topic I here just refer to a few reviews: Edwards and Palsson (1998); Kauffman *et al.* (2003); Lee *et al.* (2006).

Burgard *et al.* (2004) introduced the concept of *flux coupling analysis* that is related to the concept of FBA and yields results similar to the concepts of enzyme subsets (Pfeiffer *et al.*, 1999) or reaction correlation coefficients (Poolman *et al.*, 2007). Flux coupling analysis allows for ascertaining whether any two metabolic fluxes, V_1 and V_2, are (i) directionally coupled if a non-zero flux for V_1 implies a non-zero flux for V_2 but not necessarily the reverse; (ii) partially coupled if a non-zero flux for V_1 implicates a non-zero, but variable, flux for V_2 and vice versa; or (iii) fully coupled if a non-zero flux for V_1 entails not only a non-zero but also a fixed flux for V_2 and vice versa. Flux coupling analysis thus facilitates the global identification of blocked reactions.

1.2.5 Extreme Pathways

A concept similar to EFMs is represented by *extreme pathways* (Schilling *et al.*, 2000). As EFMs, extreme pathways are a set of convex basis vectors that is derived from the stoichiometric matrix **N**. They share the following two properties with EFMs: (i) Every given network gives rise to a unique set of extreme pathways/EFMs; (ii) Every extreme pathway/EFM consists of the minimum number of reactions that are necessary to exist as a functional unit that fulfils the steady-state condition. The third property slightly differs between the two concepts. Whereas EFMs are the set of *all* routes through

a metabolic network that are consistent with properties (i) and (ii), extreme pathways are the systemically independent subset of EFMs. This means that no extreme pathway can be represented as a non-negative linear combination of any other extreme pathways. Thus, the extreme pathways correspond to the edges of the corresponding convex polyhedral cone (as EFMs do), whereas there might be additional EFMs lying inside of the cone. The algorithms for computing EFMs and extreme pathways, respectively, handle the directionality of reactions in different ways. Extreme pathway analysis decouples all internal reversible reactions into two separate ones for the forward and the backward step, and subsequently calculates the extreme pathways. In EFM analysis the directionality of reactions is kept during the calculation process by applying a series of rules. The similarities and differences between EFM analysis and extreme pathway analysis have been extensively discussed by Papin *et al.* (2004).

An established tool for calculating extreme pathways is the `expa` (Bell and Palsson, 2005). The open-source software provides a command line interface with input options and help menu. The input file consists of an ASCII file containing either a white-space separated stoichiometric matrix or a list of metabolic reactions. The extreme pathways are returned in matrix form stored in an ASCII file. The C-code, along with binaries for `Windows`, `Linux`, and `Mac OS X`, and sample network reaction files, are available at `http://systemsbiology.ucsd.edu`.

1.2.6 Boolean Networks

Due to their application to gene-regulatory and signalling networks and for the sake of completeness, I will briefly review the concept of *Boolean networks* (BNs). BNs are based on the Boolean proposition logic that was introduced by the British mathematician and philosopher George Boole (1815–1864) and that was named after him in his honour. Boolean logic is a bivalent type of logic: any statement is either true or false. Statements can be concatenated using binary operators like "AND", "OR", or "NOT". The truth value of the resulting statement depends on the "input statements" and the way they are combined. In mathematical terms the values "true" and "false" usually corre-

spond to “1” and “0”, respectively. The logical operator “AND” is also symbolised with “$\wedge$” and the logical connection “$a \wedge b$” (also called *conjunction*) can be mapped to the arithmetic operation “$(a \cdot b)$ mod 2”. The logical operator “OR” is moreover depicted as “$\vee$” and the logical concatenation “$a \vee b$” (*disjunction*) corresponds to the arithmetic operation “$(a + b + a \cdot b)$ mod 2”. Another representation for the logical *complement* or *negation* “NOT” is “$\neg$”. In some programming languages also “!” is used. The logical statement “$\neg a$” can be translated to the arithmetic operation “$(a + 1)$ mod 2”.

Boolean logic has been applied to the analysis of biological processes such as gene-regulatory networks. The first BNs were introduced by Kauffman (1969) as random models of gene-regulatory networks. In gene-regulatory networks, genes are represented by nodes and the interactions between them correspond to the arcs connecting the nodes. The levels of gene expression are approximated by 1 and 0, meaning that a gene is either fully expressed (1) or completely not expressed (0). A network comprising m genes is hence able to assume at maximum 2^m different expression states. It depends on the interactions between the genes, which states effectively can be attained. The state values of all nodes are usually calculated in discrete time steps. The state of a successor node at time $t + 1$ depends on all states of its predecessor nodes at time t and on the kind of their interactions. There are 2^{2^k} possible Boolean rules to determine the state of a vertex with k inputs (interactions pointing to it). Boolean rules may be labeled by numbers or by representative names such as conjunction or disjunction. Usually all Boolean rules are listed in so-called *truth tables*, also called *logical tables*. Certain steady states in a system may behave as attractors, meaning that the system remains in this state once it is reached. Such attractors are called *fixed points* if they consist of just one certain state. But attractors may also occur as oscillatory behaviour if a system returns to the initial (or an intermediate) state after a certain number of steps. Let us assume a system containing three genes A, B, and C. The following two sequences of state transitions $(A_t B_t C_t) \rightarrow (A_{t+1} B_{t+1} C_{t+1})$ (two possible trajectories of the system) represent the approach to a fixed point (first trajectory) and an

oscillation with a period of 3 (second trajectory), respectively:

$$(001) \rightarrow (010) \rightarrow (100) \rightarrow (100) \tag{1.19}$$

$$(011) \rightarrow (101) \rightarrow (110) \rightarrow (011) \tag{1.20}$$

Biological clocks such as the circadian rhythms of species are assumed to be the result of complex genetic oscillations (Lakin-Thomas, 2006). Moreover, Kauffman (1969) suggested to interpret the number of possible attractors of a gene-regulatory network comprising the whole genome of a cell as the number of possible cell types the genome gives rise to.

BNs are special cases of *interaction graphs*. *Unweighted directed graphs* are defined as tupel $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ with $\mathbf{V}$ being a set of vertices (nodes) and $\mathbf{E}$ a set of edges. Furthermore holds $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$ since edges are always connecting *two* vertices (cf. Diestel, 2000). Interaction graphs (also called *causal influence graphs*) are usually *signed directed graphs* representing direct dependencies between species in biological networks (Klamt *et al.*, 2006). The vertices may have a finite number of states and the state change of a successor vertex depends on the states of the predecessor vertices (the vertices the interactions of which point to the successor vertex) and on the transition rules defined along the edges. Depending on the network type and the level of abstraction, the vertices are, for instance, genes, transcription factors, receptors, ligands, kinases, proteins, metabolites, or other compounds. Sometimes, for instance in the case of protein-protein interaction networks, interaction graphs may also be undirected since many protein-protein interactions are based on complex formations, where just the resulting complexes are relevant, but not the exact chronologies of their formation processes. Beside the directions of dependencies that are represented by arcs, the edges of interaction graphs have yet another additional property; they are labeled with a *sign* ("+" or "-") to indicate whether the influence has a positive (activating) or negative (inhibiting) effect on the state level of the successor vertex. Formally, an interaction graph is represented as a signed directed graph $\mathbf{G} = (\mathbf{V}, \mathbf{A})$ with $\mathbf{V}$ being again the set of vertices (species) and $\mathbf{A} \subseteq \mathbf{V} \times \mathbf{V} \times \{+, -\}$ being the set of labeled directed edges (arcs). A directed and signed edge (signed arc) from vertex i (tail) to vertex j (head) is

denoted by an ordered tupel (i, j, s) with $i, j \in \mathbf{V}$ and $s \in \{+, -\}$ (Klamt *et al.*, 2006). Similar to the representation of a metabolic network as a stoichiometric matrix $\mathbf{N}$, the structure of a signed interaction graph can be stored in an $m \times n$ incidence matrix $\mathbf{B}$ in which the columns correspond to the n interactions (arcs) in the system and the rows to the m species (nodes). The arc k, pointing from tail vertex i to head vertex j, is represented in the kth column of $\mathbf{B}$ with the values $\mathbf{B}_{i,k} = -1$ and $\mathbf{B}_{j,k} = +1$ (similar to substrate and product values in a stoichiometric matrix). Since arc k is only connecting the vertices i and j, all other entries in column k are zero: $\mathbf{B}_{l,k} = 0\,(l \neq i, j)$. *Self-loops* (also called *autocatalytic loops*), i.e. arcs connecting a species with itself, are not considered (Klamt *et al.*, 2006). The sign s of arc k is stored separately as the kth entry in an q-dimensional vector (cf. the sign restriction in the concept of EFMs). Similar to external metabolites in the concept of EFMs, *sources* (starting points of a signal flow) and *sinks* (end points) in signal-transduction and gene-regulatory networks can be easily identified from the incidence matrix $\mathbf{B}$ since the rows of sources and sinks contain no positive and negative entries, respectively.

A drawback of interaction graphs is the lacking possibility to concatenate several inputs via AND-connection. For instance, the interaction $A+B \rightarrow C$ (e.g. an enyzme A that can activate a kinase C not before docking to an effector B) cannot be modelled as interaction graph. Hence, Klamt *et al.* (2006) extended the interaction-graph approach to the concept of *logical interaction hypergraphs* (LIHs). Hypergraphs are generalisations of graphs. Similar to a directed graph, a *directed hypergraph* $\mathbf{H} = (\mathbf{V}, \mathbf{A_H})$ comprises a set $\mathbf{V}$ of nodes and a set $\mathbf{A_H}$ of hyperarcs (also called *directed hyperedges* (Klamt *et al.*, 2006)). Since a hyperarc connects two *subsets of nodes*, for the set of hyperarcs $\mathbf{A_H}$ holds $\mathbf{A_H} \subseteq \mathcal{P}(\mathbf{V}) \times \mathcal{P}(\mathbf{V})$ with $\mathcal{P}(\mathbf{V})$ being the power set of $\mathbf{V}$: $\mathcal{P}(\mathbf{V}) = \{\mathbf{U} : \mathbf{U} \subseteq \mathbf{V}\}$. The set of start nodes (tail vertices) $\mathbf{S}$ and the set of end nodes (head vertices) $\mathbf{E}$ of a certain hyperarc h can have arbitrary cardinality (at maximum the cardinality of $\mathbf{V}$), and a graph is a special case of a hypergraph where the cardinality of $\mathbf{S}$ and $\mathbf{E}$ is 1 for all edges (Klamt *et al.*, 2006).

In this framework, all input arcs of a node can be concatenated by the Boolean operators “AND”, “OR”, or “NOT”. The resulting Boolean function

is represented in the so-called *minimal disjunctive normal form* (also called *sum of products*), which corresponds directly to the structure of all incoming arcs. As it is the case for BNs, each species in a system is considered as a binary variable, meaning that its state is approximated by two distinct levels: "0" (= "off" or "inactive" or "absent") and "1" (= "on" or "active" or "present"). Similar to interaction graphs, LIHs can be formally represented by an $m \times n$ incidence matrix $\mathbf{B_H}$ in which the columns correspond to the n interactions (arcs) in the system and the rows to the m species (nodes). The difference to the incidence matrix $\mathbf{B}$ for interaction graphs is that in $\mathbf{B_H}$ the columns of AND-concatenated interactions are merged. Additionally to $\mathbf{B_H}$, a further matrix $\mathbf{U}$ (also of dimension $m \times n$) is necessary to cope with hyperarcs consisting of activating and inhibiting branches (e.g. in the case of an interaction like $A + \neg B \rightarrow C$, meaning that species C is only active if A is present AND B is NOT present). $U_{ik} = 1$ holds if species i is entering hyperarc k with a negated value, and $U_{ik} = 0$ otherwise (Klamt *et al.*, 2006).

Analogously to steady-state analysis in the concept of EFMs, the framework of LIHs allows for so-called *logical steady-state analysis* (LSSA). *Interaction pathways* (e.g. signalling pathways) can be computed on the basis of incidence matrices $\mathbf{B}$ or $\mathbf{B_H}$, using the same algorithms that are used for the calculation of EFMs. Nevertheless, since the program `CellNetAnalyzer` (Klamt *et al.*, 2006, 2007) does not calculate the interaction pathways on the basis of the hypergraphical structure of an interaction network but on the basis of the underlying interaction graph, and thus ignores AND-connected interactions, the interaction pathways are just partially analogous to EFMs in metabolic networks. Similar to cyclic EFMs, converting no external metabolites, in the concept of LIHs *feedback loops* are detected. Furthermore, Klamt *et al.* (2006) extended the idea of MCSs to the concept of *minimal intervention sets* (MISs). Since in signal-transduction and gene-regulatory networks also inhibitions can occur, no longer knockouts are the only way to block certain objective functions of a system. Also accessory inhibitions ("knock-ins") can achieve this intention. Hence, the framework of LIHs offers an equivalent set of tools for analysing signal-transduction and gene-regulatory networks, than the concept of EFMs does for metabolic networks. The possibility to describe signal and mass flows equivalently as interactions, allows for inte-

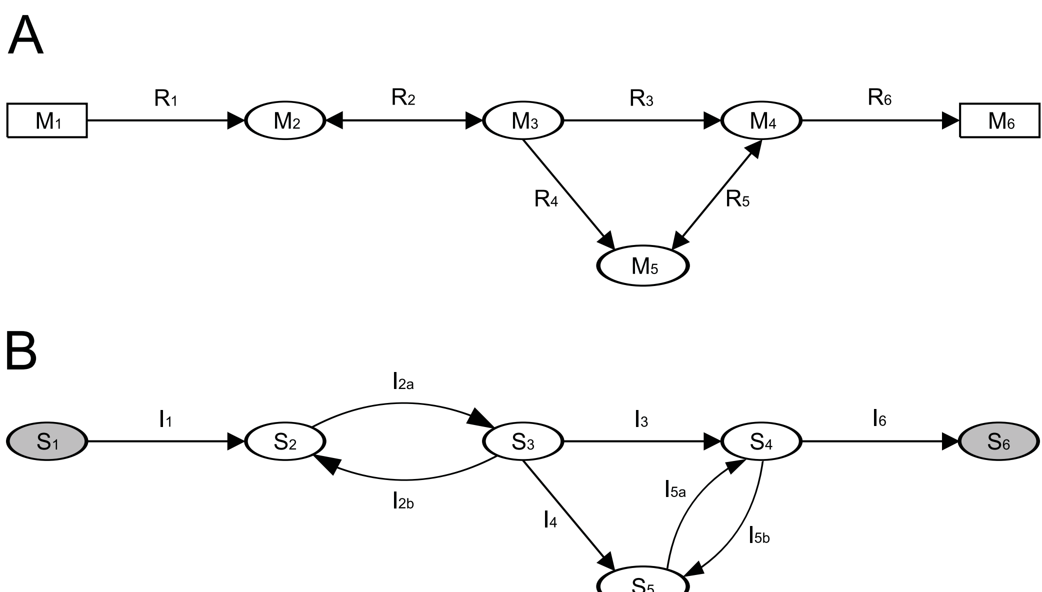

Figure 1.4. Different representations of the small prototypic example shown in Fig. 1.1. **A** Usual representation with arcs/hyperarcs. External metabolites are boxed whereas internal metabolites are depicted as ellipses. **B** Representation of the same system as an interaction (hyper)graph. Here the interactions I_1–I_6 and the species S_1–S_6 are displayed as in the usual representation. Reversible reactions are now split up into forward and backward interactions. The source species S_1 and the sink species S_6 (greyed out) are now part of the incidence matrix **B**, but can be identified since the rows of sources and sinks contain no positive and negative entries, respectively.

grating several types of cellular networks into one framework. Nevertheless, the higher level of abstraction has the disadvantage that, for instance, exact stoichiometric coefficients in mass flows can not be considered. All these tools are implemented in the program `CellNetAnalyzer` (Klamt *et al.*, 2006, 2007) that was already briefly described in Subsection 1.2.1 about the concept of EFMs.

A similar approach suggested by Gianchandani *et al.* (2006) is offering a kind of mixture of quasi-stoichiometric and Boolean concepts. Briefly described, this so-called *Matrix Formalism for Regulatory Systems* defines interactions between components of regulatory systems as quasi-stoichiometric reactions using a minimal stoichiometry being neccessary for transporting information. Participating enzymes are formally considered to be consumed and produced, respectively, as metabolites are in classical reactions, whereas on the molecular level the information flow usually consists of an inactivat-

ing modification of the former enzyme coupled with an activating modification of the following enzyme. Moreover, Gianchandani *et al.* (2006) adopt ideas from Boolean concepts by introducing "negative compounds". For each pseudo-reaction representing a regulatory rule, they additionally set up a converse reaction containing "absent metabolites". For instance, an activation of enzyme B by enzyme A is modelled as $A_{Presence} \rightarrow B_{Presence}$ and $A_{Absence} \rightarrow B_{Absence}$, whereas an inhibition of enzyme D by inhibitor C is modelled as $C_{Presence} \rightarrow D_{Absence}$ and $C_{Absence} \rightarrow D_{Presence}$. The description of a regulatory system with a quasi-stoichiometric approach gives rise to the application of all tools that work on stoichiometric matrices. Nevertheless, so far, it is not entirely clear, whether this concept is consistent with other Boolean approaches.

1.3 Robustness

Robustness is a fairly often observed property of biological systems. Since biological systems must be robust against environmental and genetic perturbations to be evolvable, robustness is a general feature of living cells that are exposed to varying conditions. These variations can be induced by the surroundings of an organism or by internal fluctuations. Soyer and Pfeiffer (2010) performed evolutionary simulations of metabolic networks under stable and fluctuating environments. They found that networks evolved under varying conditions can better tolerate single gene deletions than those networks evolved under stable conditions.

An extreme example of robustness is the anhydrobiosis (a variety of cryptobiosis) of tardigrades that suspend their metabolism almost completely under extreme dehydration and enter a dormant state that allows them to survive for years (Clegg, 2001). Kitano (2004) states that "Robustness is a property that allows a system to maintain its functions despite external and internal perturbations. It is one of the fundamental and ubiquitously observed systems-level phenomena that cannot be understood by looking at the individual components." He considers robustness to be a fundamental organisation principle of evolvable complex systems, which is plausible because evolution prefers traits enhancing the robustness of an organism. A

similar definition is used by Stelling *et al.* (2004b). Morohashi *et al.* (2002) state that "biochemical networks which are conserved across species are robust to variations in concentrations and kinetic parameters". Kitano (2007) distinguishes between homeostasis, stability, and robustness. Homeostasis and stability are related concepts, whereas robustness is a more general idea. A system is robust as long as it maintains functionality, even if it enters a new steady state to cope with perturbations. Moreover, robustness is not identical to stability. For instance, the HIV-1 virus gains robustness against many therapeutic interventions due to an increased instability concerning a high mutation rate (cf. Wagner, 2005). Hence, Kitano considers homeostasis and stability to be particular instances of robustness. He emphasises different mechanisms to ensure the robustness of a system: system control (encompassing positive and negative feedback) to achieve a robust dynamic response in a physiological range of environmental conditions), alternative mechanisms (redundancy) to compensate, for instance, the negative effects of knockouts, and modularity plus decoupling that support a higher and faster adaptability of organisms to evolutionary altering environmental conditions (Kitano, 2004).

The advantage of incorporating negative feedbacks is obvious because the need of dampening fluctuations in environmental conditions is intuitively reasonable. But the beneficial contribution of positive feedbacks to the robustness of a system seems to be contradictory to intuition having in mind that positive feedbacks usually entail bistable behaviour of a system. But ensuring clear and unique responses of a system to different stimuli needs their amplification to make them distinguishable from environmental noise.

Another important mechanism to enhance robustness in biological systems is given by redundancy meaning that there are several ways to achieve a specific function (Wagner, 2005). If, for instance, a metabolic network comprises multiple alternative routes to produce a certain metabolite, it can compensate failures of one of them by bypassing the blocked pathway. Such failures can occur due to several reasons such as genetic mutations leading to unusable or insufficiently working enzymes, but also due to environmental perturbations such as poisoning or the effects of drugs.

Redundancy is just part of a more general concept that Kitano calls *fail-*

safe mechanisms (Kitano, 2004). This concept encompasses redundancy, overlapping functions (see also Wagner, 2005) and diversity, as the differing degrees of similarity between the various alternative means that are available. To explain this in more detail let us assume that on the genetic level a gene duplication occurs leading to the existence of an isoenzyme for a certain metabolic reaction. In the course of evolution this isoenzyme can either disappear again because the loss of its (duplicated) gene is not a lethal genetic mutation. Or the gene (and thus the enzyme) may (at least slightly) change leading to an enzyme that gaines additional functions but the specificity of which still overlaps with the original one (Lewin, 2000; Deutscher *et al.*, 2006). Obviously, this example holds also for all other kinds of biological networks such as signal-transduction or gene-regulatory networks.

Getting one more step apart from the level of isoenzymes a biological network can attain its functionality also by supplying diverse pathways consisting of heterogeneous components and thus being fairly different from each other but leading to the same overall functionality. Some of these phenomena are known as *phenotypic plasticity* (Agrawal, 2001). But there are also numerous other examples of alternative mechanisms at the network level. The most often cited text-book example is probably glycolysis and oxidative phosphorylation (Berg *et al.*, 2007). Both processes produce ATP. But oxidative phosphorylation requires a constant supply of oxygen, whereas glycolysis can work either aerobic or anaerobic (although the former is more efficient). Thus, for instance, the diauxic shift in yeast causes a rigorous change in the use of different metabolic pathways, depending on whether glucose or ethanol is available for energy supply (DeRisi *et al.*, 1997). Whereas Kitano (2004) clearly distinguishes between redundancy, overlapping functions and diversity, these different aspects are usually summarised under the term redundancy.

On the metabolic level, a certain function such as the chemical conversion of a metabolite can be attained by one enzyme (that maybe consists of several amino acid chains). But, for instance, on the level of signal transduction, a specific function can be an explicit response on a certain stimulus. Such a complex process can no longer be achieved by a single component but by a concerted interplay of many components such as hormons, receptors,

kinases, adaptor proteins and many more (Kitano, 2004). On the evolutionary level the robustness of an organism is increased by encapsulating the components being necessary to maintain a certain functionality to a module that is evolutionary highly conserved. A pretty well known example of such a highly conserved modular process having a fundamental function is the transcriptional machinery of a cell, where various modules can be interfaced to create diverse phenotypes (Berg *et al.*, 2007). But also the tryptophan operon in bacteria can be considered as a module (Berg *et al.*, 2007). Introducing modules (and hence decoupling them from other modules) is also an effective mechanism for keeping the consequences of perturbations locally to minimise the damage of the whole system, and thus, it is not surprising that modules are widely observed in various organisms and often hierarchically organised. A cell, for instance is composed of various organelles, and, at the same time, it can be also part of larger modules such as tissues and organs. Modularity is even claimed to be a possible biological design principle (Kitano, 2004) but despite the intuitive consensus about the importance of the concept of modularity it is still ambivalent and therefore difficult to define (Hartwell *et al.*, 1999). On the other hand organisms gain evolutionary advantages by maintaining a certain level of adaptability. "Evolvability requires flexibility in generating diverse phenotypes by means of producing non-lethal mutations" (Hartwell *et al.*, 1999; Kirschner and Gerhart, 1998). Kirschner and Gerhart define evolutionary adaptability as a capacity to generate heritable phenotypic variations incorporating properties that "reduce the potential lethality of mutations and the number of mutations needed to produce phenotypically novel traits" (Hartwell *et al.*, 1999; Kirschner and Gerhart, 1998). To achieve this evolutionary flexibility, it is necessary to preserve functional modules as highly conserved units and to interconnect them with weak linkages. The integration of such "predetermined breaking points" serves on one hand for a higher recombinability of modules and increases on the other hand the probability that mutations affect the interconnections and not the modules. In computer science this concept of modularisation and standardised interconnections (e.g. application programming interfaces, APIs) is widely used.

The overall architecture that fits to these requirements is a nested bow-

tie structure connecting various input and output modules to a conserved core (Csete and Doyle, 2004; Kitano, 2004). At this point Kitano emphasises the similarity to *scale-free networks* (Barabási and Albert, 1999) comparing the conserved core in the bow-tie concept with the highly connected nodes in scale-free networks. These kinds of networks are able to tolerate random knockouts of nodes, but no systematic removal of nodes with high connectivity (Albert *et al.*, 2000). Jeong *et al.* (2000) and Barabási and Oltvai (2004) proposed that metabolic networks and protein-interaction networks form scale-free networks. In a scale-free network the degree distribution follows at least asymptotically a power law. Recent interest in scale-free networks started in 1999 with the work by Barabási and Albert (1999) who mapped the topology of parts of the World Wide Web, finding that some nodes, which they call *hubs*, have many more connections than others and that the network as a whole has a power-law distribution of the number of links connecting to a node. The most popular model to generate scale-free networks is the *rich get richer model* proposed by Barabási and Albert (1999), in which each new webpage creates links to existing webpages with a probability distribution that is not uniform, but proportional to the current numbers of links already pointing to these webpages. This property of scale-free networks suggests that proteins prefer to form connections to other proteins already having the highest number of links (Barabási and Oltvai, 2004; Jeong *et al.*, 2000).

Since evolution enhances robustness of organisms, one consequence following on from this is that in the course of evolution the complexity of organisms is increased through successive addition of regulatory systems. However, the introduction of various control feedback loops generates trade-offs by causing instability when unexpected perturbations are encountered, leading to severe failures (cf. Wagner, 2005). Kitano used an intuitively understandable example to illustrate these trade-offs between robustness and fragility. “The Wright brothers aeroplane is not robust against atmospheric perturbations, unlike modern commercial aeroplanes. However, modern aeroplanes are extremely fragile against unusual perturbations such as total power failure because their flight-control system is totally dependent on electricity. The Wright brothers aeroplane, on the other hand, is not affected by this type of

failure as it does not use electrical controls in the first place." (Kitano, 2004) Carlson and Doyle tried to generalise these issues in their *highly optimized tolerance theory* (HOT theory) (Carlson and Doyle, 2002) by arguing that systems that have evolved to be robust against general perturbations are extremely fragile against certain types of rare perturbations. In a more recent publication Kitano examined the question whether such trade-offs appear only in systems that are already sufficiently optimised (Kitano, 2010). In this study he predicted that the growth rate (biomass production) of organisms and cells can be enhanced in the course of evolution without an increase of fragility until a certain limit of efficiency is reached. Robustness trade-offs emerge only at this limit.

On the other hand there is an even more fundamental trade-off concerning robustness against certain perturbations and the associated metabolic effort to achieve this level of robustness (Wagner, 2005). Let us, for instance, assume that a certain pathogen developed (e.g. by mutation and selection) a resistance against an efficient antibiotic. But as long as this resistance is not crucial to survive (because the antibiotic is not yet applied) the resistance is just a metabolic overhead comsuming energy to be maintained. The original strain does not possess this resistance, and thus, it is able to use the saved energy for a more efficient reproduction. Therefore, a resistance already developed by a certain pathogen might also get lost again if the corresponding antibiotic is not dispensed for a longer time span. Just in case the antibiotic is applied, for instance, in a prophylactic manner as it is often the case in hospitals, the resistant strain will survive and propagate, leading to a return of the associated disease in a more severe form that is no longer curable with this certain antibiotic (Balaban *et al.*, 2004). Very often, a developed resistance turns on several or even a whole familiy of antibiotics (*cross-resistance*). In such cases, a completely new class of more effective antibiotics needs to be dispensed (Lüllmann *et al.*, 2006). For a review about the problem of resistant or even multi-resistant *pneumococci* in hospitals see Kristinsson (1997). See also a study by Kumarasamy *et al.* (2010) about gram-negative *Enterobacteriaceae* being resistant to the beta-lactam antibiotic "Carbapenem". There are two more examples that I want to mention in this context: firstly a study by Bloom *et al.* (2010) showing

that in H1N1 viruses (causing the 2009 flu pandemic) the single-point mutation His274 → Tyr274 that was thought to diminish the viral activity can together with several second mutations counteract the reduced viral fitness and confer oseltamivir (also known as Tamiflu) resistance to the H1N1 viruses that became predominant in the 2007–2008 flu season. Secondly, there is a study by Navarre *et al.* (2010) who reported that mutations of the two genes poxA and yjeK in *Salmonella enterica* increase the mutants' virulence and stress resistance. Due to several effects of these mutations on metabolism and transcription and translation the *Salmonella enterica* mutants become highly resistant against antibiotics and withstand many of the various effector mechanisms employed by the host immune system during infection.

Besides the way that the non-resistant strains of a bacteria colony die out whereas the resistant strains survive and grow rapidly, there exist even more complex survival strategies within such colonies. Lee *et al.* (2010) showed that even altruistic mechanisms can occur that increase the population-wide drug resistance of an *Escherichia coli* culture coming along with fitness costs for the highly resistant isolates of the population. They also revealed the mechanism behind that altruistic behaviour. The highly resistant isolates of the *Escherichia coli* culture improve the survival of the populations less resistant individuals by producing indole as a messenger substance that switches on drug efflux pumps and oxidative-stress protection mechanisms of the less resistant constituents.

Robustness analysis of biochemical networks is very often focused on variations in metabolite concentrations (Shinar and Feinberg, 2010) or on parametric sensitivities, meaning the sensitivity of systems behaviour with respect to changes in the model parameters. For example, Stelling *et al.* (2004a) studied the robustness properties of the circadian clock in *Drosophila* by systematically investigating the parameter space. Another robustness study concerning cellular rhythms was conducted by Wolf *et al.* (2005). Here, several mathematical models (of several organisms) for oscillations in calcium signalling, glycolysis, and the circadian system were compared. Jacobsen and Cedersund (2008) state that "using parametric sensitivity as a measure of robustness is in principle based on the assumption that the underlying model structure is exactly known and that all relevant perturbations can be

represented by changes in the model parameters". However, usually model structures are at least partially uncertain, for example, due to incomplete knowledge of the reaction kinetics, lumped reaction steps, and unmodelled transport phenomena, such as diffusion. Moreover, perturbations like knockouts cannot be represented by changes in the model parameters, because they are affecting the model structure itself. Thus they need to be considered by removing the whole interaction and its corresponding equation. Consequently, Jacobsen and Cedersund (2008) combine parametric and (partially) structural approaches in their study of the robustness of the oscillatory metabolism of activated neutrophils. After a classical parametric sensitivity analysis they study the change in the dynamic behaviour of the system after perturbing the direct links between the various nodes of the model.

For analysing the robustness of structural models, the above mentioned dynamic-based approaches are no longer feasible. Here, purely structural methods like EFM analysis or extreme pathways need to be applied. Edwards and Palsson (2000) analysed the metabolic capabilities of *Escherichia coli* with the help of FBA. Their analyses showed that seven reactions in the central metabolism are essential for the optimal cellular growth in glucose minimal media. The corresponding enzymes can be grouped into three categories: (1) pentose phosphate pathway, (2) three-carbon glycolytic pathway, and (3) tricarboxylic acid cycle. Deutscher *et al.* (2006) analysed the genetic robustness of the yeast metabolism by simulating multiple knockouts of genes in a flux balance model of the metabolic network in yeast. They identified gene sets providing redundant functionalities and defined the so-called *k robustness* for each gene. A system is k-robust against the knockout of a certain gene if the smallest gene set, the gene is involved in, comprises k members. Hence, in the case of an essential gene, the system is 1-robust concerning the knockout of that gene; and in the case of a *synthetic lethal* gene pair, the system is 2-robust. In a further analysis (Deutscher *et al.*, 2008) they evaluated the accuracy of single-knockout studies in finding all essential genes of a system compared with multiple-knockout studies. Thereby, they found that single-perturbations analyses miss at least 33% of the genes that contribute significantly to the growth potential of yeast. Smart *et al.* (2008) studied the robustness of metabolic networks by analysing failure cascades.

They represented the metabolic networks of *Escherichia coli*, *Methanosarcina barkeri*, *Staphylococcus aureus*, and *Saccharomyces cerevisiae* as Petri-net like bipartite graphs and analysed the networks by using topological flux balance (TFB) and thus assuming steady state. In their study they examined how perturbations caused by knockouts propagate through the networks. Due to the steady-state assumption their results are similar to those obtained by the application of EFMs or extreme pathways.

The approach of Çakir *et al.* (2004a,b) is to compare the biomass yields of wild type and mutant strains and to use these comparisons as a measure of fault-tolerance of the considered (mutant) network. Stelling *et al.* (2002) studied the structural robustness of central metabolism of *Escherichia coli in silico* using EFM analysis. In this study, they compared the wild type with different knockout mutants. If, after deletion of an enzyme, at least one EFM, allowing for a positive growth yield of biomass, remained, the mutant was considered to be viable. To get a more detailed look on the robustness, the maximum possible biomass yield was plotted versus the number of EFMs remaining in various single knockout mutants. Thereby, they could show that the biomass yield remains quite constant even if the number of EFMs is reduced significantly. This result elucidates quite well that redundancy is an important factor for the structural robustness of biological systems. But nevertheless, structural robustness is just partially equivalent with redundancy.

1.4 Outline of the thesis

The question that I address with my thesis is, how to assess the *structural robustness* of biological networks. As it is outlined in Section 1.3, robustness analyses of biochemical networks are very often focused on parametric sensitivities, meaning the sensitivity of systems behaviour with respect to changes in the model parameters (Stelling *et al.*, 2004a; Wolf *et al.*, 2005). Nevertheless, parametric sensitivity as a measure of robustness is based on the assumption that the underlying model structure is exactly known and that all relevant perturbations can be represented by changes in the model parameters (cf. Jacobsen and Cedersund, 2008). In the case that kinetic data are not or just partially available or network structures are topologi-

cally changed by knockouts, these methods are no longer feasible. In such cases structural methods need to be applied to estimate the robustness of a network in the sense of topological redundancy. It is intuitively comprehensible that a sparse network, comprising less pathways passing through it, is less robust than a highly interconnected one. However, since the pathways also need to be feasible from a biochemical point of view, it is not sufficient just to measure the interconnectivity with graph-theoretical methods (cf. de Figueiredo *et al.*, 2009b).

Hence, a measure of structural robustness needs to be developed that is based on feasible routes through biological networks and that allows for the comparison of different networks with respect to their structural fault tolerance. In defining structural robustness, one should compare the entire system with a mutated system. By considering only the number of EFMs, it remains unclear what happens to the topology of the network if an enzyme is knocked out. Thus, we introduced a more sophisticated approach to calculate the structural robustness of metabolic networks on the basis of the concept of EFMs. I present this approach (Wilhelm *et al.*, 2004) in Chapter 2 introducing the basic robustness measures.

In Chapter 3, I show the generalised approach that copes also with double and multiple knockouts (Behre *et al.*, 2008). We applied our extended concept to metabolic networks representing amino acid anabolism in *Escherichia coli* and human hepatocytes, and the central metabolism in human erythrocytes. Moreover, we subdivided the *Escherichia coli* model into two subnetworks synthesising amino acids that are essential and those that are non-essential for humans to be able to compare the amino acid anabolism of human hepatocytes with the corresponding part of *Escherichia coli*. The `Metatool` model of the amino acid anabolism of hepatocytes is listed in the Appendix of this thesis (see Appendix A.4). It comprises 82 reactions including six spontaneous reactions and—since we considered compartmentation by distinguishing between reactions in the cytosol and those in mitochondria—also 13 exchange reactions between these two compartments. During literature search I had to do for the compilation of our model, I found 25 further exchange reactions. These additional transport reactions are listed in Appendix A.5. I contributed our hepatocyte amino acid anabolism and the

additional transport reactions to the comprehensive hepatocyte metabolism *HepatoNet1* that was reconstructed in the group of Prof. Holzhütter at the Charité in Berlin (see Gille *et al.*, 2010, where I am also a co-author). Together with Jerby *et al.* (2010), HepatoNet1 is the first reconstruction of a comprehensive metabolic network of the human hepatocyte accomplishing a large number of known metabolic liver functions.

In Chapter 4, I show an extension of the concept of EFMs to signal transduction networks consisting of enzyme cascades (Behre and Schuster, 2009). Since concepts like EFMs and extreme pathways are useful tools for detecting metabolic pathways, it is tempting to adapt these methods to signalling systems. Whereas these concepts are based on a mass balance condition, in signal transduction networks it is the flow of information that matters. Here we present a formalism by which these concepts can be adapted to signal transduction networks in the case of enzyme cascades.

Chapter 2

Structural robustness concerning single knockouts

The paper of Wilhelm *et al.* (2004) provides a first introduction of our three new robustness measures based on the concept of elementary flux modes (EFMs). We show that the number of EFMs itself is not an appropriate measure of structural robustness. The robustness measures are based on the relative number of EFMs remaining after knockouts of enzymes. We discuss the relevance of these measures with the help of simple examples, as well as with larger, realistic metabolic networks. Thereby we demonstrate quantitatively that the metabolism of *Escherichia coli*, which must be able to adapt to varying conditions, is more robust than the metabolism of the human erythrocyte, which lives under much more homeostatic conditions.

Analysis of structural robustness of metabolic networks

T. Wilhelm, J. Behre and S. Schuster

Abstract: We study the structural robustness of metabolic networks on the basis of the concept of elementary flux modes. It is shown that the number of elementary modes itself is not an appropriate measure of structural robustness. Instead, we introduce three new robustness measures. These are based on the relative number of elementary modes remaining after the knockout of enzymes. We discuss the relevance of these measures with the help of simple examples, as well as with larger, realistic metabolic networks. Thereby we demonstrate quantitatively that the metabolism of *Escherichia coli*, which must be able to adapt to varying conditions, is more robust than the metabolism of the human erythrocyte, which lives under much more homeostatic conditions.

1 Introduction

A striking feature of living organisms is their homeostasis; they are, within some range, robust to external (e.g. temperature, food supply) and internal perturbations (e.g. spontaneous mutations). For example, many knockout mutants of micro-organisms are still able to grow, some with almost the same growth rate as the wild type. This has been shown, for example, by a systematic study on single knockout mutants of virtually all genes in baker's yeast [1, 2]. In many cells, there are parallel and thus redundant metabolic pathways. For example, the pentose phosphate pathway circumvents the upper part of glycolysis. Phosphoglycerate kinase in human erythrocytes can be bypassed via the Rapoport-Luebering shunt [3]. Both bypasses imply, however, a loss in ATP production. Often, redundancy in metabolism cannot be seen as easily as in these examples. To understand robustness in complex systems such as metabolic networks, theoretical tools are needed [4–8].

Studies on robustness have manifold applications. In biotechnology, specific enzymes can be knocked out to suppress futile cycles. For example, Rohwer and Botha [9] detected five futile cycles in the sucrose-accumulating sugar cane tissue. Another goal pursued in biotechnology is to suppress the synthesis of undesired products, e.g. caffeine in tea leaves [10]. It is then of interest to know which products can still be synthesised by the mutants [9, 11]. Studies on network redundancy are relevant also in medicine: How robust is human metabolism against enzyme deficiencies [12, 13]? In drug target identification, it is most suitable to find an enzyme that is non-redundant in the pathogenic micro-organism, while redundant in the human host, so that the latter is not perturbed too much. This has been analysed for the causative agent of the African sleeping disease, *Trypanosoma brucei* [14, 15].

Robustness is generally defined as the insensitivity of a system to changes in parameters [7]. These can be parameters determined by the surroundings of the organism or by internal fluctuations. Different types of robustness can be distinguished. For example, the negative feedback loops present in many biochemical pathways such as the synthesis routes of numerous amino acids make the production rate robust to changes in the demand of the product. This can be viewed as dynamic robustness. Here, however, we analyse structural robustness. We deal with the question as to whether a cell can tolerate the elimination of some enzymes by mutations. Structural robustness is necessarily linked with redundancy because the network has no other possibility in responding to a knockout than to use alternative routes.

Current methods for the modelling of metabolism have various strengths and shortcomings. Specifically, dynamic simulation of metabolic and regulatory networks [4, 14, 16, 17] meets difficulties as the necessary mechanistic detail and kinetic parameters are rarely available. In contrast, methods for analysing the topological structure of metabolic networks such as metabolic pathway analysis [11, 18–21] only require knowledge of the stoichiometric coefficients and the directionality of reactions, which is available in many cases from the literature or on-line databases.

A central concept in metabolic pathway analysis is that of elementary flux modes [18, 19]. An elementary mode is a minimal set of enzymes that can operate at steady state, such that all irreversible reactions involved are used in the appropriate direction. The enzymes are weighted by the relative flux they carry. Any flux distribution in the living cell is a superposition of elementary modes. This concept, which takes into account relevant stoichiometric and thermodynamic constraints, has been applied to a number of biochemical networks of increasing complexity (e.g. [6, 9, 22, 23]). Elementary mode analysis appears to be well-suited to characterise network redundancy because each elementary mode is non-redundant. By examining which of these modes form the same products from the same substrates, one can detect parallel routes. Using metabolic pathway analysis software, such as METATOOL [24] or FluxAnalyzer [25], this can be performed in an automated way. However, it should be mentioned that the computation

© IEE, 2004

Systems Biology online no. 20045004
doi: 10.1049/sb:20045004

T. Wilhelm is with the Institute of Molecular Biotechnology, Theoretical Systems Biology Group, Beutenbergstrasse 11, D-07745 Jena, Germany

S. Schuster is with the Friedrich Schiller University Jena, Faculty of Biology and Pharmaceutics, Section of Bioinformatics, Ernst-Abbe-Platz 2, D-07743 Jena, Germany, email: schuster@minet.uni-jena.de

J. Behre is with the Friedrich Schiller University Jena and was formerly with the Institute of Molecular Biotechnology, Jena

of elementary modes in larger networks meets the problem of combinatorial explosion [26]. Therefore, additional intellectual effort is necessary to formulate a feasible problem for a sub-network of interest.

The total number of elementary modes for given conditions has been used as a quantitative measure of network flexibility and as an estimate of fault-tolerance [6, 27]. Redundancy has also been analysed by Palsson *et al.* [20, 21]. They used the concept of extreme pathways, which is closely related to that of elementary modes. (For a comparison of the two concepts, see [28] and [29].) The simulations by Papin *et al.* [20] for *Haemophilus influenzae* showed that there was an average of 37 extreme pathways corresponding to the same input/output regime, when the network was used to produce a single amino acid. Price *et al.* [21] found by model calculations that the synthesis of amino acids and ribonucleotides in *Helicobacter pylori* is less redundant than in *H. influenzae*. In these papers, the number of extreme pathways with the same overall stoichiometry (in terms of initial substrates and final products) is used as a measure of redundancy. A slightly different approach was suggested by Oancea [30] and Çakır *et al.* [31]. They assessed the importance of each enzyme by the number of elementary modes in which it is involved or, conversely, by the number of modes remaining in the system deficient in the enzyme under study.

The structural robustness of microbial metabolism has been studied *in silico* using elementary mode analysis by Stelling *et al.* [6]. The following criterion was used: if, after deletion of an enzyme, at least one elementary mode allowing a positive growth yield remained, the mutant was predicted to be viable. To characterise robustness in more detail, the maximum possible growth yield was plotted versus the number of elementary modes remaining in various single mutants (Fig. 2*b* in [6]). This plot shows that the growth yield remains fairly constant even if the number of elementary modes drops significantly. Only when the latter number is very low, is the network no longer able to sustain growth.

In the present paper, we start from the reasoning that robustness is not perfectly identical with redundancy. In defining structural robustness, one should compare the entire system with a mutated system. By only considering the number of elementary modes, it is unclear what happens if an enzyme is knocked out. As we will show in Section 2 by way of an example, systems with the same number of elementary modes can have different robustnesses.

We propose three new measures of metabolic network robustness. They are based on elementary flux modes and take into account the effect of enzyme knockout. We illustrate the approach by simple hypothetical examples and two more complex systems describing central metabolism of *Escherichia coli* and the metabolism in human erythrocytes.

2 Measures of network robustness

Consider the two simple networks shown in Fig. 1. Both involve the same number (two) of elementary modes. However, the system in Fig. 1*a* is less robust because knockout of enzyme 1 deletes two elementary modes, while deleting only one mode in the system shown in Fig. 1*b*. Thus, the number of elementary modes (or extreme pathways) should not be used as a measure of robustness.

To characterise the structural robustness to the knockout (deficiency) of one enzyme, E_i, the ratio, $z^{(i)}/z$, between the number of elementary modes remaining after knockout $z^{(i)}$, and the number in the unperturbed network z, can be used. This gives a normalised value between 0 and 1. The extreme values are reached when no elementary mode is left (zero robustness) and when all elementary modes remain (complete robustness). This definition is actually related to the earlier suggestion to characterise the importance of each enzyme by the number of modes in which it is involved [30, 31].

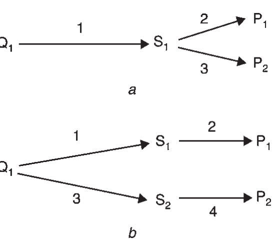

Fig. 1 *Simple hypothetical networks illustrating that the number of elementary modes is not an appropriate measure of robustness. All reactions are considered irreversible for simplicity's sake.* S_i, *internal metabolites;* Q_k, *substrates;* P_k, *products*

a The pathways producing P_1 and P_2 share one enzyme
b The pathways producing P_1 and P_2 do not share any enzyme

To quantify the global robustness of the entire network, the arithmetic mean of all these numbers can be taken

$$R_1 = \frac{\sum_{i=1}^{r} z^{(i)}}{r \cdot z} \tag{1}$$

where r denotes the total number of reactions in the system. This quantity is again between 0 and 1.

Consider the simple example shown in Fig. 2. It gives rise to four elementary modes: $\{E_1, E_2, E_4\}$, $\{E_3, E_4\}$, $\{E_5, E_6\}$, and $\{E_5, E_7\}$. Knockout of enzymes E_4 or E_5 causes two elementary modes to disappear, while knockout of one of the remaining enzymes causes only one elementary mode to drop out (that is, three to remain). Accordingly, the robustness of this system is:

$$R_1 = \frac{2 \times 2 + 5 \times 3}{7 \times 4} \cong 0.679 \tag{2}$$

This means that knockout of one enzyme implies that, on average, two-thirds of the pathways in the system are still present. For the systems shown in Figs. 1*a* and 1*b*, R_1 can readily be computed to be $1/3$ and $1/2$, respectively.

Special attention should be paid to reactions that are at thermodynamic equilibrium and hence, have zero net flux at any steady state of the system. Such reactions have been termed 'strictly detailed balanced reactions' and can be detected by analysing the nullspace matrix and checking a generalised Wegscheider condition [32]. They are not involved in any elementary mode. Therefore, their robustness measure $z^{(i)}/z$ equals one. However, for most

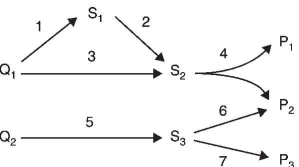

Fig. 2 *Simple hypothetical network used for illustrating the proposed robustness measures. All reactions are considered irreversible.* S_i, *internal metabolites;* Q_k, *substrates;* P_k, *products*

applications, such enzymes should be excluded from the robustness analysis because they do not contribute to any net conversion. Otherwise, the robustness of a system would change as the number of strictly detailed balanced reactions increases because their knockout does not affect the number of elementary modes while the total number of reactions r, increases. Such a dependency of robustness on strictly detailed balanced reactions is not, however, biologically meaningful. Interestingly, if they are excluded, the robustness measure can never be exactly equal to one, since knockout of an enzyme then always deletes at least one elementary mode.

In (1), no difference is made between different products of the metabolic network. For example, the contribution to the robustness measure is the same when an elementary mode producing ATP drops out and when a mode producing NADPH is eliminated. For some applications, it is certainly of interest to distinguish between different products. To cope with such situations, we propose two further definitions.

We consider the sub-network consisting of all elementary modes producing a certain essential product P_k and apply definition (1). There may be other products (excreted byproducts such as CO_2 or ethanol) which are not included in the calculation. Let us consider the example of Fig. 2. Two elementary modes, consisting of four enzymes, produce P_1. Note that both of these start from Q_1, so that they refer to the same overall stoichiometry, which would be of importance for the redundancy definition of Papin *et al.* [20]. For our definition, however, we only consider the products. The robustness concerning product P_1 is:

$$R_1^{(1)} = \frac{1+1+1+0}{4 \times 2} = 0.375 \qquad (3)$$

For the other products we obtain $R_1^{(2)} = 11/18$ and $R_1^{(3)} = 0$. The latter result is a special case of the general fact that if $R_1^{(k)} = 0$, then each enzyme of the elementary modes producing P_k is an essential enzyme as defined by Klamt and Gilles [8].

Our second measure of metabolic network robustness is based on the assumption that all of the considered products are essential for the organism. That is, the organism is no longer viable as soon as one essential product cannot be produced anymore. The measure is defined as the minimal robustness concerning an essential product:

$$R_2 = \min\left\{R_1^{(1)}, R_1^{(2)}, \ldots, R_1^{(n)}\right\} \qquad (4)$$

For the example of Fig. 2, we obtain $R_2 = 0$, because $R_1^{(3)} = 0$.

However, it may occur that the robustness of one product is quite low, but that most of the random mutations would concern elementary modes producing other, more robust products. To take this into account we propose the arithmetic mean of the particular product robustness values as a third network robustness measure

$$R_3 = \frac{\sum_i R_1^{(i)}}{n} \qquad (5)$$

with n denoting the number of essential products. The example network of Fig. 2 has the robustness $R_3 \cong 0.329$. For special purposes, it could be sensible to consider other mean values or some special weights for the particular product robustness values. Here we just cope with the simplest generic measures.

3 Realistic examples: central metabolisms of *E. coli* and human erythrocytes

As a proof of concept, we study two realistic metabolic networks that have about the same number of elementary modes. The first illustrative example is the metabolism in human erythrocytes, which is a favourite subject of modelling studies [12, 13, 16, 17, 31, 33, 34]. The network is fully described in Appendix 7.1. It is related to a model used earlier [16, 31, 33]. The choice of external metabolites and exchange reactions is similar to that used by Wiback and Palsson [34]. The scheme comprises glycolysis, the pentose phosphate pathway, glutathione oxidation/reduction and adenine nucleotide metabolism. We consider glucose, NAD, NADP, NADH, NADPH, NH_3, CO_2, and 2,3-diphosphoglycerate (2,3DPG) as external metabolites (substrates and products). Moreover, pyruvate, lactate, inosine, adenosine, adenine and hypoxanthine are considered to be exchanged across the cell membrane by reactions of the form S = S_EXT with S_EXT considered to be external. The consumption of ATP in various cell processes is modelled as a reaction ATP = ADP_EXT + Pi_EXT. Wiback and Palsson [34] considered even more of such exchange reactions (with some of them describing consumption inside the cell), which increases the number of elementary modes. Although it is questionable whether such exchange reactions are relevant in erythrocytes, we use the network in this configuration because it gives rise to about the same number of elementary modes as the *E. coli* model described below. ATP, NADPH, 2,3DPG and hypoxanthine are here considered as essential products. The functions of ATP and NADPH as energy and redox 'currencies', respectively, are well known. 2,3DPG is an effector of oxygen binding to haemoglobin [3, 17]. In addition to oxygen transport, an important function of human red blood cells is the transport of purine bases from organs with excess purine to organs in which purines are required [35]. Hypoxanthine is the relevant product excreted by erythrocytes in this context. Our present choice of external metabolites increases the number of elementary modes from 21 [33] to 667.

The second network describes the central metabolism of *E. coli* (Appendix 7.2) and is based on a model published by Stelling *et al.* [6]. The original model involves substrate uptake (including the phosphotransferase system), central carbon metabolism (including glycolysis, the pentose phosphate pathway, tricarboxylic acid cycle and glyoxylate shunt), monomer and precursor synthesis (including the synthesis of all proteinogenic amino acids and the nucleotide triphosphates), polymer synthesis, biomass production (growth) and the excretion of several byproducts such as CO_2 and formate. Further details are given in Stelling *et al.* [6] and in the on-line material mentioned therein. Since we consider four relevant products of erythrocyte metabolism, we take the same number of products in the *E. coli* model. The original model [6] involves biomass and ATP as the relevant products. Here, by way of example, we consider different combinations of four amino acids as relevant products (see Table 1). Biomass production, the remaining amino acids, as well as the other precursors of biomass are excluded for simplicity's sake. To obtain about the same number of elementary modes, we take acetate as the only substrate. Other substrates (glycerol, glucose, or succinate) would yield many more elementary modes.

The results of our calculations are shown in Table 1. For the combination alanine, arginine, asparagine, and histidine in the *E. coli* model, the number of elementary modes

Table 1: Values of the three robustness measures for the metabolic systems in human erythrocytes and *E. coli* under study

Metabolic network/ essential products	Number of elementary modes	R_1	R_2	R_3
Human erythrocyte				
ATP, hypoxanthine, NADPH, 2,3DPG	667	0.3834	0.3401	0.3607
E. coli				
Ala, Arg, Asn, His[§]	667	0.5084	0.3207	0.4295
Arg, Asn, His, Ile	656	0.5211	0.3451	0.4427
Arg, Asn, Ile, Leu	567	0.5479	0.4763	0.4964
Arg, Asn, Leu, Pro	540	0.5360	0.4586	0.4836
His, Ile, Leu, Lys	802	0.5112	0.3482	0.4437
Ile, Leu, Pro, Val	597	0.5488	0.4675	0.5058

[§] Amino acids are indicated in the usual three-letter code

exactly equals that in the erythrocyte model. Moreover, we have chosen another five combinations with about the same number of elementary modes. All the values of robustness R_1 lie between 0.38 and 0.55, implying that the number of elementary modes remaining after knockout of one enzyme is about one-third to one-half of the total number. The values of the other robustness measures are in a similar range, notably between 0.32 and 0.51. It can be seen that for all three robustness measures, the values for the erythrocyte model are lower than for the *E. coli* network, with one exception: R_2 is higher for the erythrocyte than for the combination Ala, Arg, Asn, His in *E. coli*. Apart from that, most of the values are markedly lower, especially in the case of measure R_1.

4 Conclusions

In this paper, we have proposed three new measures of network robustness. They are based on a comparison of the numbers of elementary flux modes in the unperturbed situation and after knockout of one enzyme, averaged over all enzymes.

For the measures R_2 and R_3 one has to consider all essential products separately (see Figs. 1 and 2). Note that in the general definition of elementary modes, only a distinction between internal and external metabolites is made. The latter include substrates and products. Under different conditions, one and the same external metabolite (e.g. ethanol in the case of yeast) may occur as a substrate or as a product. The robustness measures, however, should be calculated for one specific situation. In another situation, the set of essential products needs to be newly defined.

Ebenhöh and Heinrich [36] distinguish between strong and weak robustness. A metabolic network is strongly robust against a certain mutation (which in their analysis can be a knockout, change or even addition of an enzyme) if it can still produce the same products. It is weakly robust if it is not strongly robust but can still produce at least one product (not necessarily one of the original products). In our approach, a strongly robust network would have a robustness measure R_2 greater than zero (with all products taken as essential products), while for a weakly robust network, R_2 would be zero.

To illustrate the applicability of the proposed definitions, we have calculated the structural robustness measures for a model of human erythrocyte metabolism and for a model of the central metabolism of *E. coli*. The robustness of the former system is markedly lower than that of the latter system for all robustness measures and for all but one combination of products (four amino acids) of the *E. coli* system. This is in accordance with common biochemical knowledge saying that erythrocytes have a much simpler and thus, less robust metabolism than *E. coli*. The latter must be able to adapt to different conditions such as the human intestine, water of varying purity outside the human body, etc. In contrast, human erythrocytes live under relatively constant conditions in the blood. Our results are in agreement with experimental observations showing that most enzyme deficiencies, such as those of hexokinase, pyruvate kinase, or glucose-6-phosphate dehydrogenase, lead to severe diseases [37], while single-gene deletions of the majority of *E. coli* enzymes do not entail inviability. Interestingly, the calculated values of R_1 in *E. coli* ranging from 0.51 to 0.55 correspond well with the fraction 0.56 of viable mutants found by Stelling *et al.* [6] by literature search. It is worthwhile studying this relationship in more detail, both, experimentally and theoretically.

For the two considered systems, the robustness values R_1 are between 0.38 and 0.55. As most single mutants of *E. coli* are still able to grow [6], this implies that about one-third of the elementary modes in *E. coli* are often sufficient to sustain growth.

To demonstrate the applicability of our definitions, we have considered two systems with about the same number of elementary modes and the same number of biologically relevant products. This was to show that the robustness (according to our definitions) can differ nevertheless and hence, that the number of elementary modes is not an appropriate measure of robustness. However, since the proposed measures are normalised quantities, one can use them, in general, also to compare the robustness of systems with different numbers of elementary modes and different numbers of essential products. For example, we also analysed the original erythrocyte model of Joshi and Palsson [16] giving rise to 21 elementary modes [33]. We took hypoxanthine excretion and sodium/potassium transport as essential functions. This yielded the following robustness values: $R_1 = 0.4424$, $R_2 = 0.2029$, $R_3 = 0.2056$. Their difference to the values for *E. coli* is even more pronounced. In further investigations, we will apply the measures to *E. coli* under different conditions as well as to the metabolism in other cell types and other organisms. As R_2 and R_3 are defined as the minimum and average values of the same set of quantities, we obviously have $R_2 \leq R_3$. In all our numerical calculations, moreover, $R_3 < R_1$. It is worth trying to prove this relation analytically in a general way.

In future work, it will be of interest to apply the present method for comparing two micro-organisms or different groups of products in the same micro-organism, e.g. amino acid metabolism and lipid metabolism in *E. coli*. The question arises whether the present approach is scalable to larger, e.g. genome-wide, networks. This meets the above-mentioned problem of combinatorial explosion of elementary modes. However, since our robustness measures are ratios of numbers of elementary modes, it is of great interest to see whether the ratios representing the robustness measures can be computed directly without computing the elementary modes themselves. Another option is to decompose complex networks into smaller, tractable sub-networks [20, 26]. Approaches based on linear programming [5, 38] scale up more easily to larger systems. However, they only take into account the optimal situation

rather than all possible flux distributions in the system and thus, can hardly cope with network flexibility.

The definitions introduced here essentially take into account single mutants. In future work, it would be of interest to extend the analysis to double and multiple mutants. So far, there are only a few modelling studies on such mutants [38]. Klamt and Gilles [8] introduced the concept of minimal cut sets. These are minimal sets of enzymes whose deletion or complete inhibition prevents the operation of a target reaction under study. If the smallest of these sets involves, for example, two enzymes, any single mutant can still sustain the target reaction, but some double mutant cannot. Klamt and Gilles [8] introduced a fragility coefficient as the reciprocal of the average size of all minimal cut sets in which an enzyme E_i participates. A network fragility coefficient F, was defined as the average fragility coefficient over all enzymes. This coefficient takes into account both single and multiple knockouts. On the other hand, the fragility coefficient is based on an all-or-none decision whether or not a product can still be synthesised, while our measures have the advantage that the number of elementary modes is considered. Moreover, they are easier to compute. Our preliminary calculations show that for many simple systems (e.g. the systems shown in Figs. 1*a* and *b*), robustness R_1 and the coefficient F, calculated by taking all output reactions as target reactions, add up to one (note that high fragility implies low robustness), while for more complex systems such as *E. coli* and erythrocyte metabolisms, R_1 is larger than $1 - F$. For example, for the complete *E. coli* system (including biomass and ATP production) with acetate as the only substrate, our calculations give a value of $R_1 = 0.4084$, while $F = 0.783$ [8]. It will be of interest to elucidate the interrelation between these coefficients in more detail.

Further possibilities of extending the proposed definitions include the introduction of a weighted mean of product robustness values, since normally, different products are not equally important for the organism (e.g. ATP and 2,3DPG). For example, the weighting factors could be the normalised numbers of elementary modes producing the product in question or the normalised numbers of enzymes involved. The latter option appears to be sensible if each enzyme is subject to failure with the same probability.

In future studies, it will be of interest to analyse the change of robustness of metabolism during biological evolution [36]. While one would assume an increase in robustness, the opposite change can have happened as well, for example, in the evolution of intracellular parasites. In this context, it is worth studying the evolution of enzymes with broad substrate specificity. It has been argued that enzymes with high specificity have developed from low-specificity ancestors during biological evolution [39]. One reason for this development may be an increase in robustness. Two specialised enzymes cause the system to have a greater robustness than one less specific enzyme because the knockout of the latter would generally delete more pathways.

5 Acknowledgments

The authors would like to thank Steffen Klamt (Magdeburg) for stimulating discussions and the German Ministry for Education and Research for financial support to T. Wilhelm, and J. Behre. We are grateful to three anonymous referees for helpful comments.

6 References

1 Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., *et al.*: 'Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis', *Science*, 1999, **285**, pp. 901–906
2 Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., *et al.*: 'Functional profiling of the *Saccharomyces cerevisiae* genome', *Nature*, 2002, **418**, pp. 387–391
3 Bossi, D., and Giardina, B.: 'Red cell physiology', *Mol. Asp. Med.*, 1996, **17**, pp. 117–128
4 Barkai, N., and Leibler, S.: 'Robustness in simple biochemical networks', *Nature*, 1997, **387**, pp. 913–917
5 Edwards, J.S., and Palsson, B.O.: 'Robustness analysis of the *Escherichia coli* metabolic network', *Biotechnol. Prog.*, 2000, **16**, pp. 927–939
6 Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E.D.: 'Metabolic network structure determines key aspects of functionality and regulation', *Nature*, 2002, **420**, pp. 190–193
7 Morohashi, M., Winn, A.E., Borisuk, M.T., Bolouri, H., Doyle, J., and Kitano, H.: 'Robustness as a measure of plausibility in models of biochemical networks', *J. Theor. Biol.*, 2002, **216**, pp. 19–30
8 Klamt, S., and Gilles, E.D.: 'Minimal cut sets in biochemical reaction networks', *Bioinformatics*, 2004, **20**, pp. 226–234
9 Rohwer, J.M., and Botha, F.C.: 'Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data', *Biochem. J.*, 2001, **358**, pp. 437–445
10 Kato, M., Mizuno, K., Crozier, A., Fujimura, T., and Ashihara, H.: 'Caffeine synthase gene from tea leaves', *Nature*, 2000, **406**, pp. 956–957
11 Mavrovouniotis, M.L., Stephanopoulos, G., and Stephanopoulos, G.: 'Computer-aided synthesis of biochemical pathways', *Biotechnol. Bioeng.*, 1990, **36**, pp. 1119–1132
12 Schuster, R., and Holzhütter, H.G.: 'Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. Application to enzyme deficiencies of red blood cells', *Eur. J. Biochem.*, 1995, **229**, pp. 403–418
13 Martinov, M.V., Plotnikov, A.G., Vitvitsky, V.M., and Ataullakhanov, F.I.: 'Deficiencies of glycolytic enzymes as a possible cause of hemolytic anemia', *Biochim. Biophys. Acta.*, 2000, **1474**, pp. 75–87
14 Bakker, B.M., Mensonides, F.I.C., Teusink, B., Van Hoek, P., Michels, P.A.M., and Westerhoff, H.V.: 'Compartmentation protects trypanosomes from the dangerous design of glycolysis', *Proc. Natl. Acad. Sci. U.S.A.*, 2000, **97**, pp. 2087–2092
15 Wagner, C.: 'Systembiologie gegen Parasiten', *BioWorld*, 2004, **9**, (1), pp. 2–4
16 Joshi, A., and Palsson, B.O.: 'Metabolic dynamics in the human red cell. Part I. A comprehensive kinetic model', *J. Theor. Biol.*, 1989, **141**, pp. 515–528
17 Mulquiney, P.J., and Kuchel, P.W.: 'Model of 2,3-biphosphoglycerate metabolism in the human erythrocyte based on detailed enzyme kinetic equations: equations and parameter refinement', *Biochem. J.*, 1999, **342**, pp. 581–596
18 Schuster, S., and Hilgetag, C.: 'On elementary flux modes in biochemical reaction systems at steady state', *J. Biol. Syst.*, 1994, **2**, pp. 165–182
19 Schuster, S., Fell, D.A., and Dandekar, T.: 'A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks', *Nature Biotechnol.*, 2000, **18**, pp. 326–332
20 Papin, J.A., Price, N.D., Edwards, J.S., and Palsson, B.O.: 'The genome-scale metabolic extreme pathway structure in *Haemophilus influenzae* shows significant network redundancy', *J. Theor. Biol.*, 2002, **215**, pp. 67–82
21 Price, N.D., Papin, J.A., and Palsson, B.O.: 'Determination of redundancy and systems properties of the metabolic network of *Helicobacter pylori* using genome-scale extreme pathway analysis', *Genome Res.*, 2002, **12**, pp. 760–769
22 Poolman, M.G., Fell, D.A., and Raines, C.A.: 'Elementary modes analysis of photosynthate metabolism in the chloroplast stroma', *Eur. J. Biochem.*, 2003, **270**, pp. 430–439
23 Carlson, R., and Srienc, F.: 'Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Identification of reactions', *Biotechnol. Bioeng.*, 2004, **85**, pp. 1–19
24 Pfeiffer, T., Sánchez-Valdenebro, I., Nuño, J.C., Montero, F., and Schuster, S.: 'METATOOL: For studying metabolic networks', *Bioinformatics*, 1999, **15**, pp. 251–257
25 Klamt, S., Stelling, J., Ginkel, M., and Gilles, E.D.: 'FluxAnalyzer: Exploring structure, pathways, and fluxes in balanced metabolic networks by interactive flux maps', *Bioinformatics*, 2003, **19**, pp. 261–269
26 Dandekar, T., Moldenhauer, F., Bulik, S., Bertram, H., and Schuster, S.: 'A method for classifying metabolites in topological pathway analyses based on minimization of pathway number', *BioSystems*, 2003, **70**, pp. 255–270
27 Çakir, T., Kirdar, B., and Ülgen, K.Ö.: 'Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks', *Biotechnol. Bioeng.*, 2004, **86**, pp. 251–260
28 Klamt, S., and Stelling, J.: 'Two approaches for metabolic pathway analysis?', *Trends Biotechnol.*, 2003, **21**, pp. 64–69
29 Palsson, B.O., Price, N.D., and Papin, J.A.: 'Development of network-based pathway definitions: the need to analyze real metabolic networks', *Trends Biotechnol.*, 2003, **21**, pp. 195–198
30 Oancea, I.: 'Topological analysis of metabolic and regulatory networks

by decomposition methods'. PhD thesis, Humboldt University, Berlin, Germany, 2003
31 Çakir, T., Tacer, C.S., and Ülgen, K.Ö.: 'Metabolic pathway analysis of enzyme-deficient human red blood cells', *BioSystems*, 2004, in press
32 Schuster, S., and Schuster, R.: 'Detecting strictly detailed balanced subnetworks in open chemical reaction networks', *J. Math. Chem.*, 1991, **6**, pp. 17–40
33 Schuster, S., Fell, D.A., Pfeiffer, T., Dandekar, T., and Bork, P.: 'Elementary modes analysis illustrated with human red cell metabolism' in Larsson, C., Påhlman, I.-L., and Gustafsson, L., (Eds.): 'BioThermoKinetics in the Post Genomic Era' (Chalmers, Göteborg, 1998), pp. 332–339
34 Wiback, S.J., and Palsson, B.O.: 'Extreme pathway analysis of human red blood cell metabolism', *Biophys. J.*, 2002, **83**, pp. 808–818
35 Salerno, C., and Giacomello, A.: 'Hypoxanthine-guanine exchange by intact human erythrocytes', *Biochem.*, 1985, **24**, pp. 1306–1309
36 Ebenhöh, O., and Heinrich, R.: 'Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness', *Bull. Math. Biol.*, 2003, **65**, pp. 323–357
37 Scriver, C.R., and Sly, W.L., (Eds.): 'The Metabolic and Molecular Bases of Inherited Disease' (McGraw-Hill, New York, 1995), vol. III
38 Edwards, J.S., and Palsson, B.O.: 'Systems properties of the *Haemophilus influenzae* Rd metabolic genotype', *J. Biol. Chem.*, 1999, **274**, pp. 17410–17416
39 Kacser, H., and Beeby, R.: 'Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection', *J. Mol. Evol.*, 1984, **20**, pp. 38–51

7 Appendix

7.1 Input file of the erythrocyte system for the program METATOOL

The identifiers have the following meaning: ENZREV, reversible enzymes; ENZIRREV, irreversible enzymes; METINT, internal metabolites; METEXT, external metabolites; CAT, catalyzed reactions.

-ENZREV
PGI ALD TPI GAPDH PGK PGM EN LD PGL RPI XPI TKI TA TKII PRM PNPase ApK PRY:TP LAC:TP ATP:EX INO:TP ADO:TP ADE:TP

-ENZIRREV
HK PFK DPGM DPGase PK G6PDH PDGH PRPPsyn HGPRT AdPRT IMPase AMPDA AMPase ADA AK HX:TP

-METINT
G6P F6P FDP DHAP GA3P 13DPG 3PG 2PG PEP PYR LAC 6PGL 6PGC RL5P X5P R5P ADP Pi S7P E4P PRPP IMP R1P HX INO ADE ADO AMP ATP

-METEXT
NAD NADP NADH NADPH H NH3 CO2 H2O PYR_EXT LAC_EXT ADP_EXT Pi_EXT INO_EXT ADO_EXT HX_EXT ADE_EXT 23DPG

-CAT
HK : 1 ATP = 1 G6P + 1 ADP + 1 H .
PGI : 1 G6P = 1 F6P .
PFK : 1 F6P + 1 ATP = 1 FDP + 1 ADP + 1 H .
ALD : 1 FDP = 1 DHAP + 1 GA3P .
TPI : 1 DHAP = 1 GA3P .
GAPDH : 1 GA3P + 1 NAD + 1 Pi = 1 13DPG + 1 NADH + 1 H .
PGK : 1 13DPG + 1 ADP = 1 3PG + 1 ATP .
DPGM : 1 13DPG = 1 23DPG + 1 H .
DPGase : 1 23DPG + 1 H2O = 1 3PG + 1 Pi .
PGM : 1 3PG = 1 2PG .
EN : 1 2PG = 1 PEP + 1 H2O .
PK : 1 PEP + 1 ADP + 1 H = 1 PYR + 1 ATP .
LD : 1 PYR + 1 NADH + 1 H = 1 LAC + 1 NAD .
G6PDH : 1 G6P + 1 NADP = 1 6PGL + 1 NADPH + 1 H .
PGL : 1 6PGL + 1 H2O = 1 6PGC + 1 H .
PDGH : 1 6PGC + 1 NADP = 1 RL5P + 1 NADPH + 1 CO2 .
RPI : 1 RL5P = 1 R5P .
XPI : 1 RL5P = 1 X5P .
TKI : 1 X5P + 1 R5P = 1 GA3P + 1 S7P .
TA : 1 GA3P + 1 S7P = 1 F6P + 1 E4P .
TKII : 1 X5P + 1 E4P = 1 F6P + 1 GA3P .
PRPPsyn : 1 R5P + 1 ATP = 1 PRPP + 1 AMP .
PRM : 1 R1P = 1 R5P .
HGPRT : 1 PRPP + 1 HX + 1 H2O = 1 IMP + 2 Pi .
AdPRT : 1 PRPP + 1 ADE + 1 H2O = 1 AMP + 2 Pi .
PNPase : 1 INO + 1 Pi = 1 R1P + 1 HX .
IMPase : 1 IMP + 1 H2O = 1 INO + 1 H + 1 Pi .
AMPDA : 1 AMP + 1 H2O = 1 IMP + 1 NH3 .
AMPase : 1 AMP + 1 H2O = 1 ADO + 1 H + 1 Pi .
ADA : 1 ADO + 1 H2O = 1 INO + 1 NH3 .
AK : 1 ADO + 1 ATP = 1 AMP + 1 ADP .
ApK : 2 ADP = 1 AMP + 1 ATP .
PRY:TP : 1 PYR_EXT = 1 PYR .
LAC:TP : 1 LAC_EXT = 1 LAC .
ATP:EX : 1 ATP = 1 ADP_EXT + Pi_EXT .
INO:TP : 1 INO_EXT = 1 INO .
ADO:TP : 1 ADO_EXT = 1 ADO .
HX:TP : 1 HX = 1 HX_EXT .
ADE:TP : 1 ADE_EXT = 1 ADE .

7.2 Input file of the E. coli system for the program METATOOL adapted from [6]

The identifiers have the same meaning as given in Appendix 7.1. All the 20 proteinogenic amino acids are here indicated as internal metabolites. In the six different versions used in the calculations, different sets of four amino acids are set to external status (see main text).

-ENZREV
CO2_ex G6P::F6P F16P::T3P DHAP::G3P G3P::DPG DPG::3PG 3PG::2PG 2PG::PEP Cit::ICit ICit::alKG SuccCoA::Succ Fum::Mal Mal::OxA G6P::PGlac AcCoA::Adh Adh::Eth Rl5P::X5P Rl5P::R5P Transket1 Transaldo Transket2 AcCoA::AcP AcP::Ac Pyr::Lac NADHDehydro TransHydro ATPSynth MTHF_Synth

-ENZIRREV
O2_up N_up S_up DHAP::Glyc3P Lac_ex Eth_ex Ac_ex Ac_up Form_ex F16P::F6P F6P::F16P PEP::PYR Pyr::PEP PYR::AcCoA AcCoA::Cit alKG::SuccCoA Succ::Fum Fum::Succ ICit::Glyox Glyox::Mal PGlac::PGluc PGluc::Rl5P OxA::PEP PEP::OxA Pyr::Form Oxidase ATPdrain Chor_Synth PRPP_Synth Ala_Synth Val_Synth Leu_Synth Asn_Synth_1 Asp_synth Asp::Fum Asp::AspSAld AspSAld::HSer Lys_Synth Met_Synth Thr_Synth Ile_Synth His_Synth Glu_synth Gln_Synth Pro_Synth Arg_Synth Trp_Synth Tyr_Synth Phe_Synth Ser_Synth Gly_Synth Cys_Synth rATP_Synth rGTP_Synth rCTP_Synth rUTP_Synth dATP_Synth dGTP_Synth dCTP_Synth dTTP_Synth mit_FS_Synth UDPGlc_Synth CDPEth_Synth OH_myr_ac_Synth C14_0_FS_Synth CMP_KDO_Synth NDPHep_Synth TDPGlcs_Synth UDP_NAG_Synth UDP_NAM_Synth di_am_pim_Synth ADPGlc_Synth Mal::Pyr Pyr::Ac Ac::AcCoA

-METINT
G6P F6P F16P DHAP Glyc3P G3P DPG 3PG 2PG PEP Pyr AcCoA Cit ICit alKG SuccCoA Succ Fum Mal OxA Glyox R5P Rl5P E4P X5P S7P PGlac PGluc ATP NADH NADPH QuiH2 H_ex O2 CO2 N S AcP Ac Form Lac Adh Eth Chor PRPP MTHF AspSAld HSer rATP rGTP rCTP rUTP dATP dGTP dCTP dTTP mit_FS UDPGlc CDPEth OH_myr_ac C14_0_FS CMP_KDO NDPHep TDPGlcs UDP_NAG UDP_NAM di_am_pim ADPGlc Cys Asp Glu Phe Gly Ile Lys Leu Met Pro Gln Ser Thr Val Trp Tyr Ala His Asn Arg

-METEXT
O2_ext N_ext CO2_ext Lac_ext Eth_ext Ac_ext Form_ext ATP_ext

-CAT
O2_up : O2_ext = 1 O2 .
N_up : N_ext = 1 N .
CO2_ex : 1 CO2 = CO2_ext .
S_up : 4 ATP + 4 NADPH = 1 S .
DHAP::Glyc3P : 1 DHAP + 1 NADH = 1 Glyc3P .
Lac_ex : 1 Lac = Lac_ext .
Eth_ex : 1 Eth = Eth_ext .
Ac_ex : 1 Ac = Ac_ext .
Ac_up : Ac_ext = 1 Ac .
Form_ex : 1 Form = Form_ext .
G6P::F6P : 1 G6P = 1 F6P .
F16P::F6P : 1 F16P = 1 F6P .
F6P::F16P : 1 F6P + 1 ATP = 1 F16P .
F16P::T3P : 1 F16P = 1 DHAP + 1 G3P .
DHAP::G3P : 1 DHAP = 1 G3P .
G3P::DPG : 1 G3P = 1 DPG + 1 NADH .
DPG::3PG : 1 DPG = 1 3PG + 1 ATP .
3PG::2PG : 1 3PG = 1 2PG .
2PG::PEP : 1 2PG = 1 PEP .
PEP::PYR : 1 PEP = 1 Pyr + 1 ATP .
Pyr::PEP : 1 Pyr + 2 ATP = 1 PEP .
PYR::AcCoA : 1 Pyr = 1 AcCoA + 1 NADH + 1 CO2 .
AcCoA::Cit : 1 AcCoA + 1 OxA = 1 Cit .
Cit::ICit : 1 Cit = 1 ICit .
ICit::alKG : 1 ICit = 1 alKG + 1 NADPH + 1 CO2 .
alKG::SuccCoA : 1 alKG = 1 SuccCoA + 1 NADH + 1 CO2 .
SuccCoA::Succ : 1 SuccCoA = 1 Succ + 1 ATP .
Succ::Fum : 1 Succ = 1 Fum + 1 QuiH2 .
Fum::Succ : 1 Fum + 1 QuiH2 = 1 Succ .
Fum::Mal : 1 Fum = 1 Mal .
Mal::OxA : 1 Mal = 1 OxA + 1 NADH .
ICit::Glyox : 1 ICit = 1 Succ + 1 Glyox .
Glyox::Mal : 1 AcCoA + 1 Glyox = 1 Mal .
G6P::PGlac : 1 G6P = 1 PGlac + 1 NADPH .
AcCoA::Adh : 1 AcCoA + 1 NADH = 1 Adh .
Adh::Eth : 1 NADH + 1 Adh = 1 Eth .
PGlac::PGluc : 1 PGlac = 1 PGluc .
PGluc::Rl5P : 1 PGluc = 1 Rl5P + 1 NADPH + 1 CO2 .
Rl5P::X5P : 1 Rl5P = 1 X5P .
Rl5P::R5P : 1 Rl5P = 1 R5P .
Transket1 : 1 R5P + 1 X5P = 1 G3P + 1 S7P .
Transaldo : 1 G3P + 1 S7P = 1 F6P + 1 E4P .
Transket2 : 1 E4P + 1 X5P = 1 F6P + 1 G3P .
OxA::PEP : 1 OxA + 1 ATP = 1 PEP + 1 CO2 .
PEP::OxA : 1 PEP + 1 CO2 = 1 OxA .
AcCoA::AcP : 1 AcCoA = 1 AcP .
AcP::Ac : 1 AcP = 1 ATP + 1 Ac .
Pyr::Form : 1 Pyr = 1 AcCoA + 1 Form .
Pyr::Lac : 1 Pyr + 1 NADH = 1 Lac .
NADHDehydro : 1 NADH = 1 QuiH2 + 2 H_ex .
Oxidase : 1 QuiH2 + 0.5 O2 = 2 H_ex .
TransHydro : 1 NADH + 1 H_ex = 1 NADPH .
ATPSynth : 3 H_ex = 1 ATP .
ATPdrain : 1 ATP = ATP_ext .
Chor_Synth : 2 PEP + 1 E4P + 1 ATP + 1 NADPH = 1 Chor .
PRPP_Synth : 1 R5P + 2 ATP = 1 PRPP .
MTHF_Synth : 1 ATP + 1 NADPH = 1 MTHF .
Ala_Synth : 1 Pyr + 1 Glu = 1 alKG + 1 Ala .
Val_Synth : 2 Pyr + 1 NADPH + 1 Glu = 1 alKG + 1 CO2 + 1 Val .
Leu_Synth : 2 Pyr + 1 AcCoA + 1 NADPH + 1 Glu = alKG + 1 NADH + 2 CO2 + 1 Leu .
Asn_Synth_1 : 2 ATP + 1 N + 1 Asp = 1 Asn .
Asp_synth : 1 OxA + 1 Glu = 1 alKG + 1 Asp .
Asp::Fum : 1 Asp = 1 Fum + 1 N .
Asp::AspSAld : 1 ATP + 1 NADPH + 1 Asp = 1 AspSAld .
AspSAld::HSer : 1 NADPH + 1 AspSAld = 1 HSer .
Lys_Synth : 1 di_am_pim = 1 CO2 + 1 Lys .
Met_Synth : 1 SuccCoA + 1 MTHF + 1 HSer + 1 Cys = 1 Pyr + 1 Succ + 1 N + 1 Met .
Thr_Synth : 1 ATP + 1 HSer = 1 Thr .
Ile_Synth : 1 Pyr + 1 NADPH + 1 Glu + 1 Thr = 1 alKG + 1 CO2 + 1 N + 1 Ile .
His_Synth : 1 ATP + 1 PRPP + 1 Gln = 1 alKG + 2 NADH + 1 His .
Glu_synth : 1 alKG + 1 NADPH + 1 N = 1 Glu .
Gln_Synth : 1 ATP + 1 N + 1 Glu = 1 Gln .
Pro_Synth : 1 ATP + 2 NADPH + 1 Glu = 1 Pro .
Arg_Synth : 1 AcCoA + 4 ATP + 1 NADPH + 1 CO2 + 1 N + 1 Asp + 2 Glu = 1 alKG + 1 Fum+ 1 Ac + 1 Arg .
Trp_Synth : 1 Chor + 1 PRPP + 1 Gln + 1 Ser = 1 G3P + 1 Pyr + 1 CO2 + 1 Glu + 1 Trp .
Tyr_Synth : 1 Chor + 1 Glu = 1 alKG + 1 NADH + 1 CO2 + 1 Tyr .
Phe_-
Synth : 1 Chor + 1 Glu = 1 alKG + 1 CO2 + 1 Phe .
Ser_Synth : 1 3PG + 1 Glu = 1 alKG + 1 NADH + 1 Ser .
Gly_Synth : 1 Ser = 1 MTHF + 1 Gly .
Cys_Synth : 1 AcCoA + 1 S + 1 Ser = 1 Ac + 1 Cys .
rATP_Synth : 5 ATP + 1 CO2 + 1 PRPP + 2 MTHF + 2 Asp + 1 Gly + 2 Gln = 2 Fum + 1 NADPH + 2 Glu + 1 rATP .
rGTP_Synth : 6 ATP + 1 CO2 + 1 PRPP + 2 MTHF + 1 Asp + 1 Gly + 3 Gln = 2 Fum + 1 NADH+ 1 NADPH + 3 Glu + 1 rGTP .
rCTP_Synth : 1 ATP + 1 Gln + 1 rUTP = 1 Glu + 1 rCTP .
rUTP_Synth : 4 ATP + 1 N + 1 PRPP + 1 Asp = 1 NADH + 1 rUTP .
dATP_Synth : 1 NADPH + 1 rATP = 1 dATP .
dGTP_Synth : 1 NADPH + 1 rGTP = 1 dGTP .
dCTP_Synth : 1 NADPH + 1 rCTP = 1 dCTP .
dTTP_Synth : 2 NADPH + 1 MTHF + 1 rUTP = 1 dTTP .
mit_FS_Synth : 8.24 AcCoA + 7.24 ATP+ 13.91 NADPH = 1 mit_FS .
UDPGlc_Synth : 1 G6P + 1 ATP = 1 UDPGlc .
CDPEth_Synth : 1 3PG + 3 ATP + 1 NADPH + 1 N = 1 NADH + 1 CDPEth .
OH_myr_ac_Synth : 7 AcCoA + 6 ATP + 11 NADPH = 1 OH_myr_ac .
C14_0_FS_Synth : 7 AcCoA + 6 ATP + 12 NADPH = 1 C14_0_FS .
CMP_KDO_Synth : 1 PEP + 1 R5P + 2 ATP = 1 CMP_KDO .
NDPHep_Synth : 1.5 G6P + 1 ATP = 4 NADPH+ 1 NDPHep .
TDPGlcs_Synth : 1 F6P + 2 ATP + 1 N = 1 TDPGlcs .
UDP_NAG_Synth : 1 F6P + 1 AcCoA + 1 ATP + 1 Gln = 1 Glu + 1 UDP_NAG .
UDP_NAM_Synth : 1 PEP + 1 NADPH + 1 UDP_NAG = 1 UDP_NAM .
di_am_pim_Synth : 1 Pyr + 1 SuccCoA + 1 NADPH+ 1 AspSAld + 1 Glu = 1 alKG+ 1 Succ + 1 di_am_pim .
ADPGlc_Synth : 1 G6P + 1 ATP = 1 ADPGlc .
Mal::Pyr : 1 Mal = 1 Pyr + 1 NADH + 1 CO2 .
Pyr::Ac : 1 Pyr = 1 QuiH2 + 1 CO2 + 1 Ac .
Ac::AcCoA : 2 ATP + 1 Ac = 1 AcCoA .

Chapter 3

Structural robustness concerning multiple knockouts

In the paper of Behre *et al.* (2008) we present a generalised framework for analysing structural robustness of metabolic networks, based on the concept of EFMs. We extend our framework for calculating structural robustness on single knockouts (Wilhelm *et al.*, 2004). We now consider the general case of double and multiple knockouts. The robustness measures are again based on the ratio of the number of remaining EFMs after knockout vs. the number of EFMs in the unperturbed situation, but now averaged over all combinations of knockouts. We apply our extended concept to the amino acid anabolism in *Escherichia coli* and human hepatocytes, and to the central metabolism in human erythrocytes. Moreover, in the *Escherichia coli* model the two subnetworks synthesising amino acids that are essential and those that are non-essential for humans are studied individually. I contributed the hepatocyte amino acid anabolism and 25 additional transport reactions to the comprehensive hepatocyte metabolism *HepatoNet1* that was reconstructed by the group of Prof. Holzhütter at the Charité in Berlin (see Gille *et al.*, 2010, where I am also a co-author). Together with Jerby *et al.* (2010), HepatoNet1 is the first reconstruction of a comprehensive metabolic network of the human hepatocyte accomplishing a large number of known metabolic liver functions.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 252 (2008) 433–441
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Structural robustness of metabolic networks with respect to multiple knockouts

Jörn Behre[a,*], Thomas Wilhelm[b,1], Axel von Kamp[a], Eytan Ruppin[c,d], Stefan Schuster[a]

[a]*Faculty of Biology and Pharmaceutics, Section of Bioinformatics, Friedrich Schiller University Jena, Ernst-Abbe-Platz 2, D-07743 Jena, Germany*
[b]*Leibniz Institute for Age Research—Fritz Lipmann Institute, Theoretical Systems Biology Group, Beutenbergstrasse 11, D-07745 Jena, Germany*
[c]*School of Computer Sciences, Tel Aviv University, Tel Aviv 69978, Israel*
[d]*School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel*

Received 31 May 2007; received in revised form 21 September 2007; accepted 26 September 2007
Available online 9 October 2007

Abstract

We present a generalised framework for analysing structural robustness of metabolic networks, based on the concept of elementary flux modes (EFMs). Extending our earlier study on single knockouts [Wilhelm, T., Behre, J., Schuster, S., 2004. Analysis of structural robustness of metabolic networks. IEE Proc. Syst. Biol. 1(1), 114–120], we are now considering the general case of double and multiple knockouts. The robustness measures are based on the ratio of the number of remaining EFMs after knockout vs. the number of EFMs in the unperturbed situation, averaged over all combinations of knockouts. With the help of simple examples we demonstrate that consideration of multiple knockouts yields additional information going beyond single-knockout results. It is proven that the robustness score decreases as the knockout depth increases.

We apply our extended framework to metabolic networks representing amino acid anabolism in *Escherichia coli* and human hepatocytes, and the central metabolism in human erythrocytes. Moreover, in the *E. coli* model the two subnetworks synthesising amino acids that are essential and those that are non-essential for humans are studied separately. The results are discussed from an evolutionary viewpoint. We find that *E. coli* has the most robust metabolism of all the cell types studied here. Considering only the subnetwork of the synthesis of non-essential amino acids, *E. coli* and the human hepatocyte show about the same robustness.
© 2007 Elsevier Ltd. All rights reserved.

Keywords: *Escherichia coli* metabolism; Elementary flux modes; Erythrocyte metabolism; Hepatocyte metabolism; Robustness measure

1. Introduction

A general feature of living cells is their robustness to varying environmental conditions. Moreover, internal perturbations (e.g. knockout mutations or enzyme deficiencies) can be tolerated to a certain extent. Experimental and theoretical analyses of robustness have attracted increasing interest in recent years. For example, the virtual independence of biological oscillations on temperature (Ruoff et al., 2003) has been analysed. Besides such cases of dynamic robustness (Schuster and Holzhütter, 1995; Stelling et al., 2004; Wolf et al., 2005), structural robustness also has been intensively studied (Stelling et al., 2002; Çakır et al., 2004; Wilhelm et al., 2004; Lemke et al., 2004; Fong and Palsson, 2004; Klamt and Gilles, 2004; Papp et al., 2004; Blank et al., 2005; Ghim et al., 2005; Shlomi et al., 2005; Kaufman et al., 2005; Klamt, 2006; Deutscher et al., 2006). Structural robustness refers to the tolerance against changes in the structure of cellular networks. Knockout mutations and (complete) enzyme deficiencies obviously affect network structure. The analysis of structural robustness is part of the general trend of network-based approaches in which kinetic parameters are not included, motivated by the fact that kinetic parameters are often not perfectly known.

*Corresponding author. Tel.: +49 3641 949584.
E-mail addresses: jbehre@minet.uni-jena.de (J. Behre), thomas.wilhelm@bbsrc.ac.uk (T. Wilhelm), kamp@minet.uni-jena.de (A. von Kamp), ruppin@post.tau.ac.il (E. Ruppin), schuster@minet.uni-jena.de (S. Schuster).
[1]Present address. Institute of Food Research, Norwich Research Park, Colney, Norwich NR47UA, UK.

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.09.043

In a previous study we analysed the structural robustness of metabolic systems with respect to single knockouts (Wilhelm et al., 2004). We introduced various robustness measures, all of them averaging the effect of single knockouts over all enzymes in the system or all enzymes leading to a specific product. As a proof of concept, we applied these measures to metabolic networks of human erythrocytes and *Escherichia coli* central metabolism. In agreement with biochemical experience, we obtained lower robustness values for the erythrocyte network.

However, often organisms are affected by double and multiple knockouts or enzyme deficiencies. Multiple knockouts are of importance in biotechnology and medicine, for example, to suppress pathogenic bacteria. For instance, for a pair of synthetic lethal genes the knockout of two genes is fatal for the organism while a single knockout of either gene is not (Lemke et al., 2004; Schuldiner et al., 2005; Harrison et al., 2007). The most simple example is a pair of isoenzymes (Pál et al., 2005; Kuepfer et al., 2005), catalyzing the same essential reaction. A systematic experimental screening of double knockout mutants is complicated due to the large number of combinations. The most advanced screening of double mutants has been done on *Saccharomyces cerevisiae*: Tong et al. (2004) have analysed 4700 viable gene yeast deletion mutants, and Wong et al. (2004) have presented a comprehensive method to predict synthetic lethal gene pairs. First attempts have also been made for cultured human cells (Simons et al., 2001).

In the present paper, we generalise our theoretical network robustness studies by taking into account double and multiple knockouts and propose appropriate generalised robustness measures. As in our previous paper (Wilhelm et al., 2004), the analysis is based on the concept of elementary flux modes (EFMs). These are minimal sets of enzymes that can operate at steady state with all irreversible reactions used in the appropriate orientation (Schuster et al., 2000). In recent years, various biochemical systems (Stelling et al., 2002; Çakır et al., 2004; Schwender et al., 2004; Schuster and Kenanov, 2005; Krömer et al., 2006) have been studied using the concept of EFMs.

We apply the generalised robustness measures to reaction schemes of the central metabolism in human erythrocytes and amino acid synthesis in *E. coli* and human hepatocytes. The latter two systems are known to be quite redundant. Thus, it is interesting to compare their robustness. Moreover, in the *E. coli* model the two subnetworks synthesising (i) amino acids that are essential for humans and (ii) amino acids that are non-essential for humans are studied individually. The results will be discussed from an evolutionary viewpoint.

This paper is dedicated to the memory of Reinhart Heinrich, who was the highly respected academic teacher of two of the authors (S.S. and T.W.). He taught us the theoretical apparatus needed to better understand the behaviour of intracellular networks. Reinhart became famous (together with others) for establishing Metabolic Control Analysis. In that theoretical framework, less input information is necessary than in dynamic simulation. (Note that the elasticity coefficients harbour less information than the full set of kinetic parameters, cf. Heinrich and Schuster, 1996.) Later, he and his coworkers, as well as other people in the field, became interested in structural approaches, in which only a minimum input information is used, motivated by the unfortunately imperfect knowledge of kinetic parameters. Reinhart was very open to elementary-modes analysis, as witnessed by the monograph Heinrich and Schuster (1996). Besides his interest in dynamic robustness (Ruoff et al., 2003; Wolf et al., 2005), he also worked on the structural robustness of metabolism. In Ebenhöh and Heinrich (2003), a metabolic network is defined to be strongly robust against a knockout, exchange or even addition of an enzyme if it can still produce the same products. It is weakly robust if it can still produce at least one product. In the beginning of this millenium, he established the "scope" approach to elucidating the evolution of metabolic networks (Handorf et al., 2005; Ebenhöh et al., 2006). In that approach, information about the network structure and chemical formulas of substances as taken from online databases is used. Robustness issues play a role in that approach as well. Handorf et al. (2005) showed that the outcome of network expansion is in general rather robust against elimination of single or few reactions. There exist, however, crucial reactions the elimination of which leads to a dramatic reduction in the size of the network reachable in evolution.

2. Generalised measures of network robustness

Consider a metabolic network made up of a set of r enzymes $M = \{E_1, E_2, \ldots, E_r\}$. To quantify the structural robustness to the knockout (deficiency) of a subset of M, $K_i = \{E_{i,1}, E_{i,2}, \ldots\}$, the ratio, $z^{(i)}/z$, between the number of EFMs remaining after knockout, $z^{(i)}$, and the number in the unperturbed network, z, is used. The global robustness of the entire network is described by the arithmetic mean of all these numbers for all subsets K_i with the same cardinality, d:

$$R_1(d) = \sum_{i=1}^{c(d)} z^{(i)} \Big/ c(d)z \tag{1}$$

where

$$c(d) = \binom{r}{d}, \tag{2}$$

denotes the total number of these subsets (i.e., the total number of combinations of d knockouts from r reactions). $R_1(d)$ is the robustness with respect to knockout of exactly d enzymes. It is the average number of remaining EFMs divided by the total number of EFMs in the original system. Thus, Eq. (1) can also be written as $R_1(d) = \langle z^{(i)}(d) \rangle / z$. Since each term $z^{(i)}$ in the numerator

of Eq. (1) is as most as large as z, it follows that the quantity $R_1(d)$ is between 0 and 1.

In the special case $d = 1$, the measure $R_1(1)$ coincides with the robustness measure for single knockouts defined earlier (Wilhelm et al., 2004). Moreover, $R_1(r) = 0$ because knocking out all enzymes in a system (of enzyme-catalysed reactions) obviously deletes all EFMs. Generally, the following inequality holds

$$1 = R_1(0) \geqslant R_1(1) \geqslant R_1(2) \geqslant R_1(3) \geqslant \ldots R_1(r-1) \geqslant R_1(r) = 0. \quad (3)$$

This can be rationalised as follows. As mentioned above, the robustness measure is the average number of EFMs remaining after knockout of d enzymes, divided by the total number of EFMs in the original system. For $d = k+1$, we have

$$R_1(k+1) = \frac{\langle z^{(j)}(k+1) \rangle}{z}. \quad (4)$$

Since the number of EFMs cannot increase as the number of out-knocked enzymes increases, we have

$$z^{(i)}(k) \geqslant z^{(j)}(k+1) \text{ for all } i, j \text{ with } K_i(k) \subset K_j(k+1). \quad (5)$$

This implies that also the average number cannot increase, which leads to relation (3). A more detailed proof is given in Appendix A.1.

The measures $R_1(d)$ are defined separately for different knockout depths. It is tempting to define an overall measure by combining them in an appropriate way. We define this in the following way:

$$R_1(\leqslant D) = \sum_{d=1}^{D} R_1(d) p_d \text{ with } D \leqslant r, \quad (6)$$

where the p_d are weighting factors for a knockout of d enzymes together. We focus on the situation where at least one and at most D enzymes have been knocked out. In this case p_d is related to the conditional probability that the knockout which occurred is of depth d, and hence has to fulfill the normalisation condition

$$\sum_{d=1}^{D} p_d = 1. \quad (7)$$

Except for the more thorough analysis in Fig. 1, in this paper we choose $D = 3$ or 5, for two reasons: first, it is computationally too demanding to consider all knockout-combinations for a medium number of deleted enzymes ($d \sim r/2$, cf. Eq. (2)), and secondly, the higher robustness measures $R_1(d > 5)$ usually become very small and therefore do not make an important contribution to the overall measure (see Fig. 1).

The choice of appropriate weighting factors p_d is motivated by two facts: First, multiple knockouts or enzyme deficiencies occur less and less frequently as the knockout depth increases. Second, the number of possible knockout combinations changes according to $\binom{r}{d}$. Taken together this gives the weighting factors

$$p_d = p^d \binom{r}{d}. \quad (8)$$

That means, the factor p should then be calculated by solving the equation:

$$\sum_{d=1}^{D} p^d \binom{r}{d} = 1. \quad (9)$$

To get an idea of the values of the weighting factors we present, in Table 1, some exemplifying p_d values. Interestingly, the p_d converge quickly to finite values in the limit of very large reaction networks.

Moreover, we now generalise our previously defined robustness measures for specific products (Wilhelm et al., 2004) to the case of multiple knockouts. To this end, we consider the subnetwork consisting of all elementary modes producing a certain essential product P_k and apply definition (1), giving $R_1^{(k)}$. If all products of the network are essential, that is, if the mutant is not viable as soon as one product cannot be produced anymore, we define

$$R_2(d) = \min\left\{ R_1^{(1)}(d), R_1^{(2)}(d), \ldots, R_1^{(n)}(d) \right\}, \quad (10)$$

where the superscript refers to the index of the essential product. More generally, no single product might be

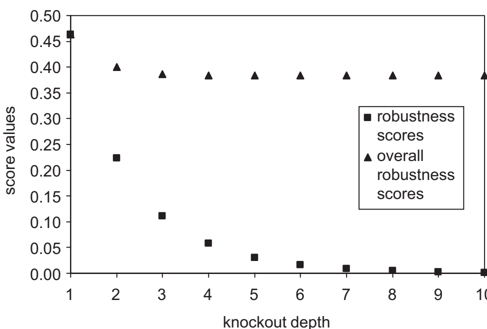

Fig. 1. Plot of the robustness score $R_1(d)$ (squares) and overall robustness score $R_1(\leqslant D)$ (triangles) vs. knockout depth (d resp. D) for the erythrocyte model. The plotted values are listed in the Supplementary Material.

Table 1
Weighting factors for $D = 3$ and the r values used in Table 2 and for $r \to \infty$

r	p_1	p_2	p_3
4	0.757	0.215	0.027
5	0.745	0.222	0.033
8	0.727	0.231	0.042
$r \to \infty$	0.699	0.244	0.057

essential, while a combination of different products is. In such a case each superscript indicates a unique smallest set of essential products.

In contrast, if no particular product (or set of products) is absolutely essential, the average of the individual product robustnesses is another appropriate robustness measure:

$$R_3(d) = \frac{\sum_{i=1}^{n} R_1^{(i)}(d)}{n}. \tag{11}$$

We have here defined the measures $R_2(d)$ and $R_3(d)$ for the sake of completeness. However, we will not further use them because we concentrate on a general comparison of networks without focussing on single products.

3. Simple examples

To point out differences in robustness to single and multiple knockouts, we consider the simple examples given in Table 2. A comparison of systems 1 and 2 shows that the measures corresponding to multiple knockouts carry additional information: although both systems have the same number of substrates, products, internal metabolites, reactions, EFMs, and the same $R_1(1)$, system 1 is less robust with respect to double and triple knockouts

$$\text{Network 1}: \quad R_1(2) = \frac{1+0+0+0+0+1}{6 \times 2} = \frac{1}{6},$$

$$\text{Network 2}: \quad R_1(2) = \frac{1+1+0+1+0+0}{6 \times 2} = \frac{1}{4}.$$

In both cases, the numbers in the numerators correspond to the following double knockouts: {1,2}, {1,3}, {1,4}, {2,3}, {2,4}, {3,4}. This means that after knockout of two enzymes, on average, one-sixth and one-fourth of the pathways in systems 1 and 2 are still present. The higher robustness of network 2 against double knockouts can intuitively be understood: the probability that double knockouts affect the same branch and, hence, leave the other branch unperturbed, is higher in network 2 than in network 1. System 2 is more robust than system 1 also with

Table 2
Simple examples of reaction schemes demonstrating different features of the robustness measures[a]

Example system	Number of reactions	Number of elementary modes	$R_1(1)$	$R_1(2)$	$R_1(3)$	$R_1(\leqslant 3)$
1	4	2	$1/2 = 0.5$	$1/6 \approx 0.167$	0	0.414
2	4	2	$1/2 = 0.5$	$1/4 = 0.25$	$1/8 = 0.125$	0.436
3	4	2	$3/8 = 0.375$	$1/12 \approx 0.083$	0	0.302
4	4	2	$1/4 = 0.25$	0	0	0.189
5	8	2	$7/16 \approx 0.438$	$3/8 = 0.375$	$5/16 \approx 0.313$	0.418
6	8	2	$1/2 = 0.5$	$3/14 \approx 0.214$	$1/14 \approx 0.071$	0.416
7	5	4	$13/20 = 0.65$	$3/8 = 0.375$	$7/40 = 0.175$	0.573
8	5	3	$2/3 \approx 0.667$	$2/5 = 0.4$	$1/5 = 0.2$	0.592

[a]For the definition of robustness measures, see text. The shown reaction networks involve monomolecular reactions only, and the metabolites at the upper and lower ends of reaction chains are defined external.

respect to triple knockouts: $R_1(3)$ of system 1 is zero, because no EFM remains, regardless of the chosen three out-knocked enzymes, whereas system 2 still has a triple-knockout robustness of $R_1(3) = 1/8$. For comparison, the networks 3 and 4 are even less robust than network 1. It can be seen that for branched systems with parallel pathways the location of the branching point is important. In system 2, which diverges at the upper end, the robustness is positive even for triple knockouts, while a branching point near the lower end can lead to a zero robustness for double and triple knockouts (system 4).

Interestingly, system 6 has a higher robustness with respect to single knockouts, but a lower double- and triple-knockout robustness than system 5. As can be seen by comparing examples 3 and 5 in Table 2, a lumping of enzymes does affect the robustness values. A lumping of all consecutive enzymes in the upper branch (which is an enzyme subset in the terminology of Pfeiffer et al., 1999) leads to the same system in both cases, notably a system consisting of three (super)enzymes and one internal metabolite acting as a branch point. Thus, the robustness values are equalised for the two systems, whereas the values for the original systems are different.

Systems 7 and 8 (both with five reactions) demonstrate that the pure number of EFMs is not an appropriate robustness measure, because system 7 has more EFMs (four), but has nevertheless a lower robustness than system 8 (three EFMs).

4. Robustness of central metabolisms in *E. coli* and human

First we analyse the central metabolism of human erythrocytes, using the network model of Schuster et al. (1998) comprising $n = 36$ internal metabolites, $r = 41$ reactions, and giving rise to 21 EFMs. The calculation procedure is explained in Appendix A.2. The EFMs are computed by the program METATOOL 5.0 (von Kamp and Schuster, 2006). The calculated robustness measures are given in Table 3. As expected, relation (3) is fulfilled for this system. Note that $R_1(\leqslant 3)$ is slightly higher than $R_1(\leqslant 5)$. This is understandable because the system is less robust when the knockout depth is larger. Mathematically, it is due to a change in the weighting factor p when D changes in the normalisation condition (9).

In Fig. 1 the robustness scores are plotted up to $d = 10$. It can be seen that knockout depths $d \geqslant 5$ are practically negligible. The overall robustness hardly changes even for $D \geqslant 3$, due to the monotonically decreasing weighting factors.

Second, we have compiled the amino acid synthesis network of *E. coli*. The reaction equations and the information about their reversibilities were taken from the databases KEGG (http://www.genome.jp/kegg/) and EcoCyc (http://www.ecocyc.org/). In total, the network contains 164 reactions (involving one spontaneous reaction) and 119 internal metabolites. The list of reaction equations including the names of external metabolites is given in the Supplementary Material. We then subdivided that network into two subnetworks: the system comprising the synthesis of amino acids which are essential for humans ($n = 89$, $r = 111$) and the system producing non-essential amino acids ($n = 53$, $r = 81$ comprising one spontaneous reaction). Note that there is a considerable overlap between the two subsystems, notably glycolysis, TCA cycle, etc. Although tyrosine is indicated as non-essential in most textbooks because it can be synthesised from phenylalanine in one step, we have here classified it as essential (in agreement with Voet and Voet (2004)) because phenylalanine is essential. The set of EFMs obtained involves several cycles of transaminase reactions. We have eliminated these because they are thermodynamically infeasible. The whole system gives rise to 65,836 EFMs (excluding such cycles), while for the subnetworks corresponding to essential and non-essential amino acids, 6874 and 11,435 EFMs, respectively, are calculated. Interestingly, the number is higher for the non-essential amino acids although the reaction number is lower (81 versus 111 for the essential amino acids). This is due to the higher degree of interconnectivity.

Table 3
Robustness measures for single and multiple knockouts up to $d = 5$ for the erythrocyte model of Schuster et al. (1998)

d (number of out-knocked enzymes)	$R_1(d)$
1	0.463
2	0.223
3	0.112
4	0.058
5	0.031
$\leqslant 3$	0.386
$\leqslant 5$	0.383

In the knockout studies, we took care that in the case of multifunctional enzymes, all reactions catalysed by a given enzyme are knocked out simultaneously. Spontaneous reactions cannot be knocked out because they are not catalysed by enzymes which could be inhibited. Therefore we skipped them during the calculations of the robustness measures. Table 4 shows the robustness values for the *E. coli* networks. For the subsystems we calculated the robustnesses up to $d = 5$. However, for the entire system, we limited the calculations to $d = 3$ due to memory restrictions. Since biologically, multiple knockouts with $d > 3$ are very rare, we do not lose important information in this way (see also Fig. 1). It can also be seen in Tables 2–5 that the overall robustness measures for the cases $d \leqslant 3$ and $d \leqslant 5$ are similar. As expected, the robustness of the *E. coli* metabolism is higher than that of erythrocyte metabolism (Tables 3 and 4). Moreover, the subnetwork corresponding to non-essential amino acids has a higher robustness at each knockout level than the network producing the essential amino acids. Furthermore, the entire network has robustness values that are above the maximum values

Table 4
Robustness measures for single and multiple knockouts up to $d = 5$ for the amino acid synthesis network of *E. coli*

d (number of out-knocked enzymes)	$R_1(d)$
Entire system	
1	0.776
2	0.602
3	0.468
$\leqslant 3$	0.716
Non-essential amino acids	
1	0.654
2	0.430
3	0.285
4	0.191
5	0.129
$\leqslant 3$	0.579
$\leqslant 5$	0.576
Essential amino acids	
1	0.623
2	0.397
3	0.258
4	0.172
5	0.117
$\leqslant 3$	0.548
$\leqslant 5$	0.544

Table 5
Robustness measures for single and multiple knockouts up to $d = 5$ for the amino acid synthesis network of human hepatocyte

Number, d, of out-knocked enzymes	$R_1(d)$
Non-essential amino acids	
1	0.659
2	0.443
3	0.305
4	0.214
5	0.154
$\leqslant 3$	0.587
$\leqslant 5$	0.584

of the two subnetworks. This is because additional routes starting and ending in one subnetwork and passing the other subnetwork drop out upon decomposition of the network. If tyrosine is taken as non-essential, similar robustness values are obtained (e.g. 0.701 and 0.616 with $d = 1$ for non-essential and essential amino acids, respectively).

Third we analysed the amino acid synthesis network of human hepatocytes (producing only non-essential amino acids, of course). The reaction equations were taken from KEGG (http://www.genome.jp/kegg/) and (mainly for the information about their reversibilities) from HumanCyc (http://www.humancyc.org/). In total, the network contains 82 reactions (involving six spontaneous reactions) and 59 internal metabolites. We considered compartmentation by distinguishing between reactions proceeding in the cytosol and those in mitochondria. We included the exchange reactions between these two compartments. The list of reaction equations including the names of external metabolites is given in the Supplementary Material. Spontaneous (non-enzymatic) reactions, multifunctional enzymes and futile cycles (again cycles of transaminase reactions) are handled as in the *E. coli* networks. For this system we calculated 712 EFMs.

Table 5 shows the robustness values for the hepatocyte network. Like the *E. coli* subnetworks, this system is, on average, still working even after the knockout of five enzymes. As expected, the hepatocyte metabolism is more robust than the erythrocyte metabolism (Tables 3 and 5), because of the minimalist and almost non-redundant metabolism of erythrocytes. Interestingly, the subnetwork corresponding to non-essential amino acids in *E. coli* and the amino acid anabolism in human hepatocytes have nearly the same robustness.

5. Discussion

Here, we have presented generalised robustness measures for metabolic networks, taking into account single, double and multiple knockouts. These measures are based on the ratio of the number of EFMs in the unperturbed situation vs. the number of remaining EFMs after knockout of one enzyme (cf. Wilhelm et al., 2004) or several enzymes, averaged over all combinations of knockouts. Since this is a normalised quantity, it does not depend on the size of the network. Moreover it is only based on network topology, so that the robustness values are the same for different metabolic systems having the same topology. With the help of simple examples (Table 2), we have demonstrated that consideration of double and triple knockouts yields additional information beyond the single-knockout studies.

We have proven that the robustness decreases if the cardinality of knockouts, d, increases. The overall robustness is, for all systems considered, between the robustness against single knockouts and that against double knockouts (Tables 3–5). It would be interesting to prove this in a general way. This observation is likely to be related with the facts that both the robustness measures (cf. Eq. (3)) and the weighting factors in the overall robustness represent monotonically decreasing series.

The examples show that for given numbers of enzymes and EFMs the robustness against double or multiple knockouts is higher if two branches in the network have different lengths (see systems 1 and 2 in Table 2). Interestingly, this is often the case in metabolism. For example, the different amino acids are synthesised on pathways of very different lengths. Although this probably has mainly chemical reasons, it might be that robustness issues also played a role in metabolic network evolution. We were not able to devise two example systems for which $R_1(3)$ is different, although $R_1(1)$ and $R_1(2)$ are equal for

the two systems. It is an interesting question whether or not two such reaction schemes can be found.

To illustrate the applicability of the new concepts, we have analysed networks of the central metabolisms in *E. coli* and in human erythrocytes and hepatocytes. Among these, the erythrocyte model shows the lowest robustness. In contrast to our previous work (Wilhelm et al., 2004) where enzymes were lumped to some extent, here we have considered each enzyme separately. As can be seen by comparing examples 3 and 5 in Table 2, a lumping of enzymes does affect the robustness values. This holds true even if the combined enzymes operate in fixed flux proportions due to structural constraints (enzyme subsets according to the definition in Pfeiffer et al., 1999).

As expected, hepatocyte metabolism is more robust than erythrocyte metabolism. This is because erythrocytes must be as small as possible in order to pass thin capillaries and are densely packed with haemoglobin for oxygen transport. Therefore only the most necessary enzymes have been retained in evolution. Both erythrocytes and hepatocytes are living under relatively homoeostatic conditions with hepatocytes having much more metabolic capabilities. In contrast, *E. coli* must adapt to widely varying situations. Thus it needs to be even more robust than hepatocytes. As a consequence *E. coli* synthesises all amino acids while hepatocytes (as all human cells) can save the metabolic effort for producing those amino acids being essential for human. A comparison of the hepatocyte network with the corresponding subnetwork of *E. coli* (just the non-essential amino acids) shows, interestingly, slightly higher robustnesses for the hepatocyte. However, the entire amino acid network of *E. coli* is significantly more robust. One reason is that the compartmentation in hepatocytes implies transporters forming bottlenecks in the system. Since not every metabolite can cross intracellular membranes, it can be hypothesised that compartmentation reduces structural robustness in many cell types.

Our analysis of the *E. coli* networks shows that amino acids essential for humans are less robustly produced than the non-essential amino acids. It is tempting to speculate that this might be the reason why their synthesis pathways got lost in the evolution towards higher organisms such as humans. The structural background for this difference is that the synthesis pathways of essential amino acids (such as tryptophan or isoleucine) are relatively ''straight'', that is, they do not involve many branch points. The enzyme genes corresponding to some of these pathways are gathered in operons (e.g. Trp operon), so that a mutational loss of the whole pathway occurs easily. In contrast, the synthesis of the non-essential amino acids runs on pathways with a higher degree of ramification and is embedded in the entangled synthesis network of other compounds, giving rise to much more redundancy.

Our approach is important for future applications in pharmacology and biotechnology. Combination drugs of two or more enzyme inhibitors have recently attracted increasing interest, while the progress in detecting drugs acting on single proteins has slowed down (Huang, 2001; Frantz, 2005). Combination therapies are of interest in treating bacterial infections (Barchiesi et al., 2004), AIDS (Taburet et al., 2004) and others. Similarly, in biotechnology, when inefficient pathways are to be suppressed, often undesired side reactions need to be deleted as well, so that multiple knockouts are necessary.

In metabolic modelling, isoenzymes are often lumped into combined reactions. While this is appropriate for many applications, it is not when robustness to enzyme deletions is studied because particular enzymes rather than particular reactions are knocked out. Thus, for example, succinate dehydrogenase and fumarate reductase in *E. coli* need to be distinguished. Analogously, enzymes with broad substrate specificity require special attention. The knockout of an enzyme catalyzing several reactions implies the deletion of all these reactions (unless they are catalysed by other enzymes simultaneously).

Some of the previous studies on robustness have tackled the question what percentage of enzymes is essential. Estimates range from about 20% (cf. Papp et al., 2004) to 30% (cf. Blank et al., 2005). The question is difficult to answer, though, because variations in external conditions are hard to take into account. Anyway, the percentage of essential enzymes is not the only relevant robustness measure. The knockout of a non-essential enzyme can have widely different effects, depending on which other enzymes are knocked out simultaneously. We have here made an attempt to quantify these effects by the average number of the remaining EFMs. A more detailed approach would consider not only the average but the diversity of effects: Is the knockout of some pair of enzymes lethal to the system and the knockout of another pair completely irrelevant, or is the effect always moderate?

The relationship of our approach to the concept of minimal cut sets (Klamt and Gilles, 2004; Klamt, 2006) is worth discussing. Minimal cut sets are minimal sets of enzymes whose suppression prevents a target reaction under study from operating. When such a set includes one reaction only, this reaction is obviously essential. When the smallest minimal cut set involves, for example, two enzymes, the target reaction can still proceed in any single-knockout mutant while it cannot in the double knockout case corresponding to that cut set. Klamt and Gilles (2004) introduced a fragility coefficient for each enzyme as the reciprocal of the mean size of all minimal cut sets in which this enzyme is included. A network fragility coefficient, F, was defined by averaging over all enzymes, taking into account both single and multiple knockouts. The fragility coefficient is based on whether or not a desired substance can still be produced, while our measures take into account the number of feasible synthesis routes. The mathematical relationship between the two concepts is not straightforward and an interesting subject of future studies. Another interesting challenge is to compare the robustness of the subnetworks of amino acids recruited early in evolution and of amino acids accrued

later (Jordan et al., 2005). It can be presumed that early adopted amino acids are more cross-linked in the network and thus lead to higher robustness.

Moreover it should be possible to extend our concept of structural robustness to gene regulatory and signal transduction networks. In that context it would be interesting to calculate the robustness of the network motifs studied by Alon (2007). Another interesting question is the extensibility of metabolic networks by additional reactions ("knock-ins"). However for this it would be necessary to define a set of plausible additional reactions.

Acknowledgements

Financial support by the BMBF (German Ministry for Education and Research) to J. Behre (HepatoSys Program) and T. Wilhelm (Jena Centre for Bioinformatics) and from the German-Israeli Foundation to A. von Kamp and E. Ruppin is gratefully acknowledged. We are grateful to Sabrina Hoffmann (Charité, Berlin) for providing information about the hepatocyte reaction network.

Appendix A.1. Proof of relation (3)

Let us compare the situations where $d = k$ with that where $d = k+1$. The sum in Eq. (1) may involve a different number of terms in these two situations. To relate these sums, we write each term of the sum in Eq. (1) for the case $d = k$ as a sum of $r-k$ identical terms divided by $r-k$:

$$\sum_{i=1}^{c(k)} z^{(i)} = \frac{\sum_{i=1}^{c(k)}\sum_{j=1}^{r-k} z^{(i)}}{r-k}. \tag{A.1}$$

The decomposition of each term of the sum in Eq. (1) corresponds to the transition from a subset K_i for $d = k$ to the $r-k$ situations of $k+1$ knockouts in which k knockouts are the same as in K_i. For example, when $d = 1$, the term describing the knockout of enzyme 1 corresponds to the terms describing the knockout of enzymes {1,2}, {1,3}, etc. in the case $d = 2$. In the set of knockout combinations thus generated, each K_i (for $d = k+1$) occurs $k+1$ times because it can come from $k+1$ original situations. For example, the term describing the knockout of enzymes {1,2} corresponds to the terms describing the knockout of enzyme 1 and to the knockout of enzyme 2 in the case $d = 1$. Therefore, we can combine the terms of the sum in Eq. (A.1) such that

$$R_1(k) \geqslant \frac{(k+1)\sum_{j=1}^{c(k+1)} \tilde{z}^{(j)}(k)}{(r-k)c(k)z} \quad \text{with} \quad K_i(k) \subset K_j(k+1) \tag{A.2}$$

$\tilde{z}^{(j)}(k)$ denotes the minimum of the remaining EFMs for each of the corresponding knockout combinations of k enzymes. For instance, $\tilde{z}^{(1,2,3)}(2) = \min\{z^{(1,2)}(2), z^{(1,3)}(2), z^{(2,3)}(2)\}$. Making use of the relation

$$\binom{r}{k+1} = \binom{r}{k}\frac{r-k}{k+1}, \tag{A.3}$$

we obtain

$$R_1(k) \geqslant \frac{\sum_{j=1}^{c(k+1)} \tilde{z}^{(j)}(k)}{c(k+1)z}. \tag{A.4}$$

This sum has the same number of terms as the sum describing $R_1(k+1)$. Therefore, it can be compared, term by term, with that sum. Since the number of EFMs cannot increase as the number of out-knocked enzymes increases, we have

$$\tilde{z}^{(i)}(k) \geqslant z^{(j)}(k+1) \text{ for all } i,j \text{ with } K_i(k) \subset K_j(k+1). \tag{A.5}$$

for each term. This leads to relation (3).

Appendix A.2. Calculation of reaction robustnesses

In principle the calculation of the robustness measures is straightforward. However, for d-knockouts in a network with r reactions, $\binom{r}{d}$ combinations (cf. Eq. (2)) have to be calculated. The number of combinations therefore increases drastically with the knockout depth d. In order to keep the calculation practical, the knockout combinations are not calculated on the level of reactions but on the level of enzyme subsets (Pfeiffer et al., 1999) and from these results the robustness is determined. The advantage is that the number of enzyme subsets is often significantly smaller than the number of reactions in the network. In order to calculate $R_1(d)$ the first step is to determine the number of remaining modes for all subset-knockout combinations of all depths from 1 to d. The next step consists of summing the number of remaining modes for depths $k = 1,\ldots, d$ whereby each subset-knockout combination at depth k is weighted with the number of different ways it can be knocked out by knocking out exactly d reactions from that subset combination (e.g. a subset combination with two times two reactions can be knocked out in four ways when $d = 2$). Note that these weights depend on d. This means that if the subset-knockout combinations of all depths from 1 to d have been calculated, the robustnesses $R_1(1),\ldots, R_1(d)$ can be computed from them by summing over them with different weighting factors. Finally, the sums calculated in this way have to be properly normalised in order to obtain the robustness values.

Appendix B. Supplementary materials

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.jtbi.2007.09.043.

References

Alon, U., 2007. An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman & Hall/CRC, Boca Raton, FL.

Barchiesi, F., Spreghini, E., Schimizzi, A.M., Maracci, M., Giannini, D., Carle, F., Scalise, G., 2004. Posaconazole and amphotericin B combination therapy against *Cryptococcus neoformans* infection. Antimicrob. Agents Chemother. 48 (9), 3312–3316.

Blank, L.M., Kuepfer, L., Sauer, U., 2005. Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol. 6 (R49), 1–16.

Çakır, T., Tacer, C.S., Ülgen, K.Ö., 2004. Metabolic pathway analysis of enzyme-deficient human red blood cells. Biosystems 78 (1–3), 49–67.

Deutscher, D., Meilijson, I., Kupiec, M., Ruppin, E., 2006. Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat. Genet. 38 (9), 993–998.

Ebenhöh, O., Heinrich, R., 2003. Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness. Bull. Math. Biol. 65 (2), 323–357.

Ebenhöh, O., Handorf, T., Kahn, D., 2006. Evolutionary changes of metabolic networks and their biosynthetic capacities. IEE Proc. Syst. Biol. 153 (5), 354–358.

Fong, S.S., Palsson, B.Ø., 2004. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nat. Genet. 36 (10), 1056–1058.

Frantz, S., 2005. Drug discovery: playing dirty. Nature 437 (7061), 942–944.

Ghim, C.-M., Goh, K.-I., Kahng, B., 2005. Lethality and synthetic lethality in the genome-wide metabolic network of *Escherichia coli*. J. Theor. Biol. 237 (4), 401–411.

Handorf, T., Ebenhöh, O., Heinrich, R., 2005. Expanding metabolic networks: scopes of compounds, robustness, and evolution. J. Mol. Evol. 61 (4), 498–512.

Harrison, R., Papp, B., Pál, C., Oliver, S.G., Delneri, D., 2007. Plasticity of genetic interactions in metabolic networks of yeast. Proc. Natl Acad. Sci. 104 (7), 2307–2312.

Heinrich, R., Schuster, S., 1996. The Regulation of Cellular Systems. Chapman & Hall, New York.

Huang, S., 2001. Genomics, complexity and drug discovery: insights from Boolean network models of cellular regulation. Pharmacogenomics 2 (3), 203–221.

Jordan, I.K., Kondrashov, F.A., Adzhubei, I.A., Wolf, Y.I., Koonin, E.V., Kondrashov, A.S., Sunyaev, S., 2005. A universal trend of amino acid gain and loss in protein evolution. Nature 433 (7026), 633–638.

Kaufman, A., Keinan, A., Meilijson, I., Kupiec, M., Ruppin, E., 2005. Quantitative analysis of genetic and neuronal multi-perturbation experiments. PLoS Comput. Biol. 1 (6), 500–506.

Klamt, S., 2006. Generalized concept of minimal cut sets in biochemical networks. Biosystems 83 (2–3), 233–247.

Klamt, S., Gilles, E.D., 2004. Minimal cut sets in biochemical reaction networks. Bioinformatics 20 (2), 226–234.

Krömer, J.O., Wittmann, C., Schröder, H., Heinzle, E., 2006. Metabolic pathway analysis for rational design of L-methionine production by *Escherichia coli* and *Corynebacterium glutamicum*. Metab. Eng. 8, 353–369.

Kuepfer, L., Sauer, U., Blank, L.M., 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. Genome Res. 15 (10), 1421–1430.

Lemke, N., Herédia, F., Barcellos, C.K., dos Reis, A.N., Mombach, J.C.M., 2004. Essentiality and damage in metabolic networks. Bioinformatics 20 (1), 115–119.

Pál, C., Papp, B., Lercher, M.J., 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. 37 (12), 1372–1375.

Papp, B., Pál, C., Hurst, L.D., 2004. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429 (6992), 661–664.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuño, J.C., Montero, F., Schuster, S., 1999. METATOOL: for studying metabolic networks. Bioinformatics 15 (3), 251–257.

Ruoff, P., Christensen, M.K., Wolf, J., Heinrich, R., 2003. Temperature dependency and temperature compensation in a model of yeast glycolytic oscillations. Biophys. Chem. 106 (2), 179–192.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., Weissman, J.S., Krogan, N.J., 2005. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123 (3), 507–519.

Schuster, R., Holzhütter, H.G., 1995. Use of mathematical models for predicting the metabolic effect of large-scale enzyme activity alterations. Application to enzyme deficiencies of red blood cells. Eur. J. Biochem. 229 (2), 403–418.

Schuster, S., Kenanov, D., 2005. Adenine and adenosine salvage pathways in erythrocytes and the role of *S*-adenosylhomocysteine hydrolase—a theoretical study using elementary flux modes. FEBS J. 272 (20), 5278–5290.

Schuster, S., Fell, D.A., Pfeiffer, T., Dandekar, T., Bork, P., 1998. Elementary modes analysis illustrated with human red cell metabolism. In: Larsson, C., Påhlman, I.-L., Gustafsson, L. (Eds.), BioThermoKinetics in the Post Genomic Era. Chalmers, Göteborg, pp. 332–339.

Schuster, S., Fell, D.A., Dandekar, T., 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. Nat. Biotechnol. 18 (3), 326–332.

Schwender, J., Goffman, F., Ohlrogge, J.B., Shachar-Hill, Y., 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. Nature 432 (7018), 779–782.

Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes following genetic perturbations. Proc. Natl Acad. Sci. 102 (21), 7695–7700.

Simons, A., Dafni, N., Dotan, I., Oron, Y., Canaani, D., 2001. Establishment of a chemical synthetic lethality screen in cultured human cells. Genome Res. 11 (2), 266–273.

Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., Gilles, E.D., 2002. Metabolic network structure determines key aspects of functionality and regulation. Nature 420 (6912), 190–193.

Stelling, J., Sauer, U., Szallasi, Z., Doyle 3rd, F.J., Doyle, J., 2004. Robustness of cellular functions. Cell 118 (6), 675–685.

Taburet, A.M., Raguin, G., Le Tiec, C., Droz, C., Barrail, A., Vincent, I., Morand-Joubert, L., Chêne, G., Clavel, F., Girard, P.-M., 2004. Interactions between amprenavir and the lopinavir–ritonavir combination in heavily pretreated patients infected with human immunodeficiency virus. Clin. Pharmacol. Ther. 75 (4), 310–323.

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., Chen, Y., Cheng, X., Chua, G., Friesen, H., Goldberg, D.S., Haynes, J., Humphries, C., He, G., Hussein, S., Ke, L., Krogan, N., Li, Z., Levinson, J.N., Lu, H., Ménard, P., Munyana, C., Parsons, A.B., Ryan, O., Tonikian, R., Roberts, T., Sdicu, A.-M., Shapiro, J., Sheikh, B., Suter, B., Wong, S.L., Zhang, L.V., Zhu, H., Burd, C.G., Munro, S., Sander, C., Rine, J., Greenblatt, J., Peter, M., Bretscher, A., Bell, G., Roth, F.P., Brown, G.W., Andrews, B., Bussey, H., Boone, C., 2004. Global mapping of the yeast genetic interaction network. Science 303 (5659), 808–813.

Voet, D., Voet, J.G., 2004. Biochemistry. Wiley, Hoboken, New Jersey.

von Kamp, A., Schuster, S., 2006. Metatool 5.0: fast and flexible elementary modes analysis. Bioinformatics 22 (15), 1930–1931.

Wilhelm, T., Behre, J., Schuster, S., 2004. Analysis of structural robustness of metabolic networks. IEE Proc. Syst. Biol. 1 (1), 114–120.

Wolf, J., Becker-Weimann, S., Heinrich, R., 2005. Analysing the robustness of cellular rhythms. IEE Proc. Syst. Biol. 2 (1), 35–41.

Wong, S.L., Zhang, L.V., Tong, A.H.Y., Li, Z., Goldberg, D.S., King, O.D., Lesage, G., Vidal, M., Andrews, B., Bussey, H., Boone, C., Roth, F.P., 2004. Combining biological networks to predict genetic interactions. Proc. Natl Acad. Sci. 101 (44), 15,682–15,687.

Chapter 4

Modelling signal transduction in enzyme cascades

In the paper of Behre and Schuster (2009) we extend the concept of EFMs to signal transduction networks consisting of enzyme cascades. Since concepts like EFMs and extreme pathways are useful tools for detecting metabolic pathways, it is tempting to adapt these methods to signalling systems. In metabolic networks these concepts are based on the compliance of a mass balance condition. In signal transduction networks this condition is of minor importance because it is the flow of information that matters. Here, we present a formalism by which these concepts can be adapted to signal transduction networks in the case of enzyme cascades operating, for example, by phosphorylation and dephosphorylation. Our approach is based on the ideas that the signal is not diminished along each route and that the system has to return to its original state after each signalling event. We illustrate the method by several simple prototypic single-phosphorylation and double-phosphorylation cascades including convergent and divergent branching. Moreover, we apply it to a specific example from insulin signalling.

JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 16, Number 6, 2009
© Mary Ann Liebert, Inc.
Pp. 829–844
DOI: 10.1089/cmb.2008.0177

Modeling Signal Transduction in Enzyme Cascades with the Concept of Elementary Flux Modes

JÖRN BEHRE and STEFAN SCHUSTER

ABSTRACT

Concepts such as elementary flux modes (EFMs) and extreme pathways are useful tools in the detection of non-decomposable routes (metabolic pathways) in biochemical networks. These methods are based on the fact that metabolic networks obey a mass balance condition. In signal transduction networks, that condition is of minor importance because it is the flow of information that matters. Nevertheless, it would be interesting to apply pathway detection methods to signaling systems. Here, we present a formalism by which this can be achieved in the case of enzyme cascades operating, for example, by phosphorylation and dephosphorylation. It is based on the ideas that the signal is not diminished along each route and that the system has to return to its original state after each signaling event. We illustrate the method by several simple prototypic single-phosphorylation and double-phosphorylation cascades, including convergent and divergent branching. Moreover, it is applied to a specific example from insulin signaling. (See online Supplementary Material at *www.liebertonline.com.*)

Key words: double phosphorylation, elementary flux modes, insulin signaling network, kinase cascades, signal transduction networks.

1. INTRODUCTION

Signal transduction is essentially important for all biological functions in living cells. The modeling of signaling networks has recently attracted considerable interest (Steffen et al., 2002; Papin and Palsson, 2004; Xiong et al., 2004; Zevedei-Oancea and Schuster, 2005; Klamt et al., 2006; Sackmann et al., 2007; Arga et al., 2007). In metabolic networks, mass flow is a crucial aspect, while in signal transduction networks, it is the flow of information that matters. Furthermore, metabolic networks usually subsist in steady states, a fact that is used in the modeling of these networks (Savageau, 1976; Heinrich and Schuster, 1996; Schilling et al., 2000; Klamt, 2006). In contrast, in signaling systems, time-dependent signals (e.g., short pulses) occur. Several approaches to model structural properties of signal transduction networks have been presented, for example, by Papin and Palsson (2004) and Xiong et al. (2004). These analyses used the concept of extreme pathways, introduced by Schilling et al. (2000) to model metabolic networks at steady states. This method is based on the earlier concept of extreme currents (Clarke, 1981). A further method for

Faculty of Biology and Pharmaceutics, Section of Bioinformatics, Friedrich Schiller University Jena, Jena, Germany.

modeling signaling systems is based on the concept of minimal T-invariants, which had been introduced in the theory of Petri nets, first for non-biological, technological applications (Starke, 1990). Several studies using this concept have been made by Heiner et al. (2004) and Sackmann et al. (2006, 2007). Extreme pathways and minimal T-invariants are similar to the concept of elementary flux modes (Schuster and Hilgetag, 1994; Schuster et al., 1999). The latter concept has widely been used in the structural modeling of metabolic networks (Stelling et al., 2002; Schwender et al., 2004; Schwartz et al., 2007; Behre et al., 2008). A comparison between minimal T-invariants and elementary flux modes and, in general, the modeling by Petri nets with traditional modeling in biochemistry has been given by Zevedei-Oancea and Schuster (2003), and a comparison between extreme pathways and elementary flux modes, by Papin et al. (2004).

Another approach for modeling signal transduction networks are logical interaction hypergraphs, which were introduced by Klamt et al. (2006). This approach reduces the processes on the molecular level to a Boolean perspective. A limitation of that method is that reactions with more than one product are difficult to model.

General difficulties in modeling signal transduction networks arise because the flow of information does not necessarily have the same direction as the mass flow that is required to transport the information. In kinase cascades, for example, the mass flow consists of cycles of enzymes being phosphorylated and dephosphorylated while the information is transduced through the cascade. In some other systems, the phosphate moiety is transferred along the reaction chain, for example, in the phosphotransferase system (PTS). In that case, the flow of information corresponds to the flow of mass, so that methods from metabolic modeling can be used in a straightforward way as shown by Papin and Palsson (2004) using a prototypic example. In the vast majority of signaling systems, however, mass flow does not correspond to information flow. Nevertheless, several authors have applied these methods formally to such systems (Xiong et al., 2004; Heiner et al., 2004; Sackmann et al., 2006, 2007). This has led to meaningful results, although the theoretical foundation for this approach is not really clear.

The present article deals with the theoretical justification for applying methods of structural analysis of metabolic systems to signal transduction networks composed of enzyme cascades. We propose an explanation based on the reasoning that (i) such networks usually allow signal amplification, that is, the signal amplitude must at least remain constant, and (ii) that, after the signal has been transmitted, the system must regain its original state so that a steady state can be assumed in the sense of an average over longer time spans. When discussing enzyme cascades, we take phosphorylation as a prototypic covalent modification. Nevertheless, any other form of chemical modification switching on or off the activity of the enzyme—such as methylation, acetylation, adenylylation, or ubiquitination—can be described by the same formalism. Even phosphorylation can occur in various ways depending on which nucleotide phosphate is used. Here, we write NTP and NDP as unspecified symbols for nucleotide triphosphates and diphosphates, with ATP and ADP being the most common representatives.

2. ASSUMING STEADY STATE FOR SIGNAL TRANSDUCTION NETWORKS

As mentioned above, in metabolic networks, mass flow is the important aspect, while in signal transduction networks, the flow of information is more important. Of course, in signaling networks, mass flow exists as well, but its direction is not necessarily the same as that of the information flow. This is particularly obvious in kinase cascades. Upon entry of an initial signal, the first kinase E_1 of the cascade is phosphorylated (leading to E_1P). Let us assume that it thus becomes active, which is indeed the most common case. Now it catalyzes the phosphorylation of the next kinase and so on. After some time, E_1P will become inactive again by dephosphorylation. Each phosphorylated kinase can catalyze the phosphorylation of the next kinase several times without being consumed. Therefore, there is no mass flow from one kinase to the next while information is transferred. This can even lead to amplification of the signal.

An important assumption made for modeling metabolic networks is the steady-state condition. It is crucial for calculating elementary flux modes (Schuster and Hilgetag, 1994), extreme pathways (Schilling et al., 2000), and minimal T-invariants (Heiner et al., 2004). In signaling systems, by contrast, time-dependent signals occur, for example, short pulses. Nevertheless, there is an important fact that allows applying the steady-state condition also to signaling systems: these systems must be regenerated after each signaling event. Let us consider, for example, the above-mentioned signaling cascade again. Because of the dephosphorylation after the pulse, the concentration of E_1 returns to its initial state. Averaged over the time

period that is needed for phosphorylation and dephosphorylation of E_1, its concentration stays constant during repeated phosphorylation/dephosphorylation cycles.

On the other hand, the steady state is just an approximation rather than a strict condition. Also in metabolic networks, fluctuations of concentrations occur (e.g., due to diffusion), but they are usually neglected because they average out (Segel, 1993; Heinrich and Schuster, 1996). Therefore, sometimes the term "pseudo-steady-state" is used.

In line of the above reasoning, it can be argued that, for detecting potential routes on which pulses can be propagated across the network, it is sufficient to detect the routes on which a flow can subsist at steady state. This can be understood by a hydrodynamic analog. A flood wave on a river can usually go only along the course on which that river flows at constant water level. If one or more rivers form a network by confluence of rivers or divergence in estuaries, the flood waves can go on the various branches of the network. An exception is when the wave is so high that the water flows over the river bank. Then, new routes can be used. This is, in signaling networks, somewhat analogous to a flow along side branches that are normally at equilibrium.

3. APPLYING THE CONCEPT OF ELEMENTARY FLUX MODES TO ENZYME CASCADES

Motivated by the above reasoning, we here apply the concept of elementary flux modes (EFMs) to enzyme cascades.

A mathematically detailed definition of EFMs was given by Schuster et al. (2002). Here, we briefly recapitulate their most important aspects for the sake of better understanding. An EFM is a minimal set of enzymes that can operate at steady state with all irreversible reactions used in the appropriate direction. All flux distributions in the living cell are non-negative linear combinations of EFMs.

The steady-state condition is given by the equation system

$$NV(S) = 0 \tag{1}$$

with N being the stoichiometric matrix and $V(S)$ being the vector containing all fluxes in the system. Together, with the sign restriction for irreversible fluxes

$$V^{irr} \geq 0 \tag{2}$$

this gives a linear equation/inequality system. For any flux vector V with elements v_k, let

$$S(V) = \{i : v_i = 0\} \tag{3}$$

Each EFM V^* fullfills conditions (1) and (2) as well as the following non-decomposability condition. There exists no vector V^{**} (unequal to the null vector) that obeys conditions (1) and (2), and contains zero components wherever V^* does and in at least one additional position,

$$S(V^*) \subset S(V^{**}) \tag{4}$$

The solution of equation system (1, 2) is a convex polyhedral cone in flux space. The edges of this cone correspond to EFMs. Additional EFMs may lie inside this cone (Schuster et al., 2002).

For computing EFMs, we use the program Metatool, developed earlier in our group (von Kamp and Schuster, 2006).

3.1. Single-phosphorylation cascades

When the phosphorylated form of some enzyme, E_iP, is to convey a signal to the next level, it must catalyze the phosphorylation of at least one molecule of E_{i+1}, and can then be dephosphorylated. Of course, it can also catalyze the phosphorylation of more than one molecule of E_{i+1}. Then, amplification occurs. The reaction equation can be written as

$$E_iP + n\ E_{i+1} + n\ NTP = E_i + P + n\ E_{i+1}P + n\ NDP \tag{5}$$

Here, we have combined the phosphorylation of E_{i+1} and the subsequent dephosphorylation of E_i into one reaction. We have $n \geq 1$, because diminuition of the signal is irrelevant. The minimum is phosphorylation of one molecule. Considering, for the moment, only this lower limit, we can write

$$E_iP + E_{i+1} + NTP = E_i + P + E_{i+1}P + NDP \tag{6}$$

This reaction equation applies to all levels i of the cascade. The application of this reaction equation to a single-phosphorylation cascade comprising the kinases E_1–E_5 gives rise to one EFM, as expected. Its overall reaction is:

$$E_1P + E_5 + 4\ NTP = E_1 + E_5P + 4\ NDP + 4\ P \tag{7}$$

The corresponding Metatool input and output is given in the Supplementary Material (see online supplementary material at *www.liebertonline.com*).

An important case is where some enzyme E_i can phosphorylate two different kinases, F_1 and G_1, so that a branching in the cascade occurs. Then, there are two minimum situations: either E_iP activates exactly one F_1 or exactly one G_1 before being dephosphorylated. Thus, reactions (8a,b) occur:

$$E_iP + F_1 + NTP = E_i + P + F_1P + NDP \tag{8a}$$

$$E_iP + G_1 + NTP = E_i + P + G_1P + NDP \tag{8b}$$

We have modeled a diverging cascade comprising the kinases E_1–E_3, F_1–F_3, and G_1–G_3. The cascade is shown in Figure 1 with reactions (8a,b) depicted by dash-dotted arrows. The corresponding Metatool input is given in Table 1. This cascade contains an OR-connection at the branching point because the phosphorylations of F_1 and G_1 are not necessarily coupled (see eqs. 8a,b).

As one would expect for an OR-connection at the branching point, this model comprises two EFMs. The first one consists of the reactions R_1, R_2, R_3, R_4, and R_5, the second one of R_1, R_2, R_6, R_7, and R_8. These reaction lists already indicate the routes of information flow. The corresponding overall reactions are

$$E_1P + F_3 + 5\ NTP = E_1 + F_3P + 5\ NDP + 5\ P \tag{9a}$$

for the first EFM, and for the second one

$$E_1P + G_3 + 5\ NTP = E_1 + G_3P + 5\ NDP + 5\ P \tag{9b}$$

TABLE 1. METATOOL INPUT FOR THE OR-CONNECTED DIVERGING CASCADE SHOWN IN FIGURE 1

```
-ENZREV

-ENZIRREV

R1 R2 R3 R4 R5 R6 R7 R8

-METINT
E2 E2P E3 E3P F1 F1P F2 F2P G1 G1P G2 G2P

-METEXT
NDP NTP E1 E1P F3 F3P G3 G3P P

-CAT
R1 : 1 NTP + 1 E2 + 1 E1P = 1 E2P + 1 NDP + 1 E1 + 1 P
R2 : 1 NTP + 1 E3 + 1 E2P = 1 E3P + 1 NDP + 1 E2 + 1 P
R3 : 1 NTP + 1 F1 + 1 E3P = 1 F1P + 1 NDP + 1 E3 + 1 P
R4 : 1 NTP + 1 F2 + 1 F1P = 1 F2P + 1 NDP + 1 F1 + 1 P
R5 : 1 NTP + 1 F3 + 1 F2P = 1 F3P + 1 NDP + 1 F2 + 1 P
R6 : 1 NTP + 1 G1 + 1 E3P = 1 G1P + 1 NDP + 1 E3 + 1 P
R7 : 1 NTP + 1 G2 + 1 G1P = 1 G2P + 1 NDP + 1 G1 + 1 P
R8 : 1 NTP + 1 G3 + 1 G2P = 1 G3P + 1 NDP + 1 G2 + 1 P
```

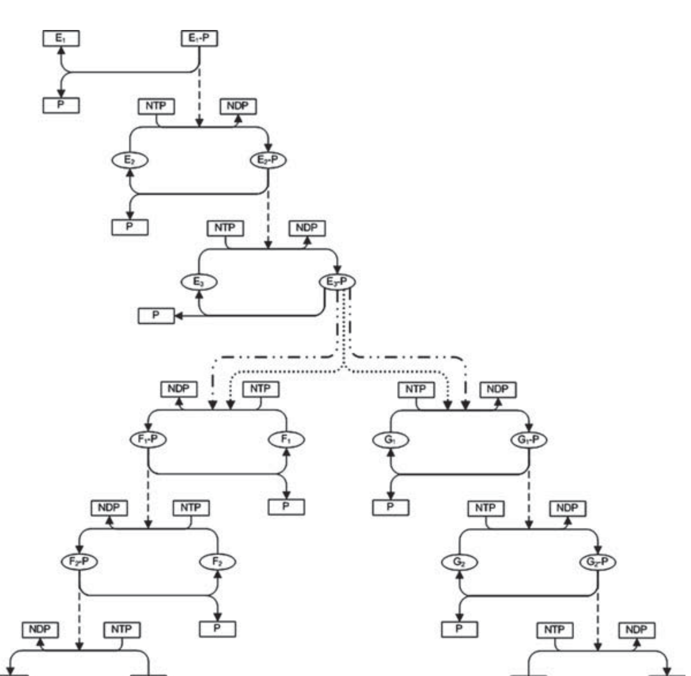

FIG. 1. Diverging cascade, comprising the kinases E_1–E_3, F_1–F_3, and G_1–G_3. External and internal metabolites are represented by rectangles and ellipses, respectively. The OR-connection is depicted by dash-dotted arrows, and the AND-connection is depicted by dotted arrows.

The same routes are relevant when signal amplification occurs, that is, when $n > 1$. Even if one molecule of E_1 catalyzes the phosphorylation of more than one molecule of F_1 and of more than one molecule of G_1, then the signal flow from E_1 to F_3 and G_3 can be decomposed into two elementary routes: from E_1 to F_3 and from E_1 to G_3.

The complete Metatool output is given in the Supplementary Material (see online supplementary material at *www.liebertonline.com*). In the following, we will show some selected Metatool inputs in the text. All other inputs and outputs are given in the Supplementary Material (see online supplementary material at *www.liebertonline.com*).

In the case that E_3P can catalyze the phosphorylation of F_1 only when G_1 is also phosphorylated, the diverging cascade is AND-connected. For this case, we modified the model above by merging the reactions (8a) and (8b) to the new reaction:

$$2\ NTP + F_1 + G_1 + E_3P = F_1P + G_1P + 2\ NDP + E_3 + P \tag{10}$$

The cascade is shown in Figure 1 with reaction (10) depicted by a dotted arrow.

As expected, only one EFM is calculated. It consists of all reactions in the model, and its overall reaction reads as follows:

$$E_1P + F_3 + G_3 + 8\ NTP = E_1 + F_3P + G_3P + 8\ NDP + 7\ P \tag{11}$$

Also, for converging cascades, one can distinguish between OR-connection and AND-connection. Thus, we also model these two variants.

The converging cascade with OR-connection is shown in Figure 2 with reactions (12a,b) depicted by dash-dotted arrows. The corresponding Metatool input is given in Table 2. The OR-connection is given by the two reactions:

$$NTP + G_1 + E_3P = G_1P + NDP + E_3 + P \tag{12a}$$

$$NTP + G_1 + F_3P = G_1P + NDP + F_3 + P \tag{12b}$$

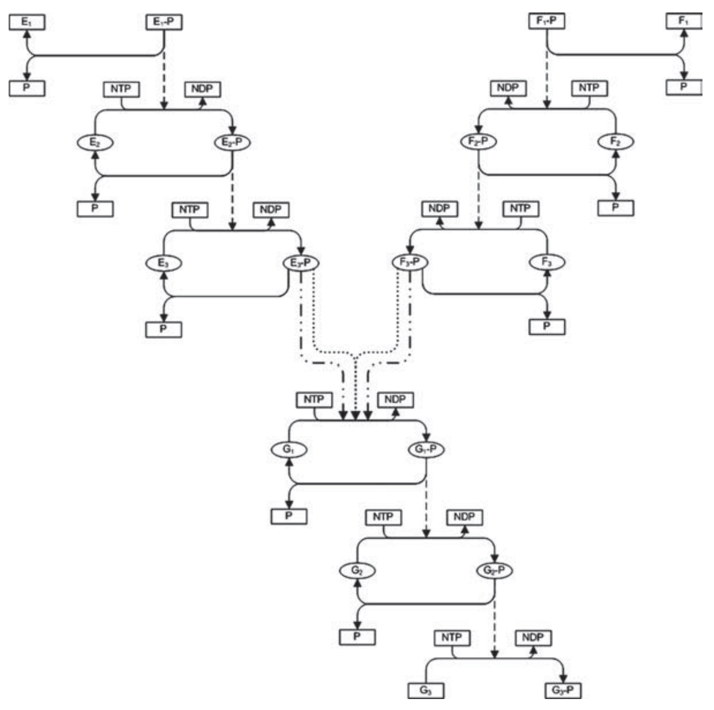

FIG. 2. Converging cascade, comprising the kinases E_1–E_3, F_1–F_3, and G_1–G_3. Rectangles, external metabolites; ellipses, internal metabolites; dash-dotted arrows, OR-connection; dotted arrows, AND-connection.

Again, the OR-connection gives rise to two EFMs, as expected. The first EFM consists of the reactions R_1, R_2, R_5, R_7, and R_8, and the second one of R_3, R_4, R_6, R_7, and R_8. The corresponding overall reactions are

$$E_1P + G_3 + 5\ NTP = E_1 + G_3P + 5\ NDP + 5\ P \tag{13a}$$

for the first EFM, and for the second one

$$F_1P + G_3 + 5\ NTP = F_1 + G_3P + 5\ NDP + 5\ P \tag{13b}$$

The converging cascade with AND-connection is shown in Figure 2 with reaction (14) depicted by a dotted arrow. Analogously to the diverging cascade with AND-connection, here also the branching point is represented by one reaction:

$$NTP + G_1 + E_3P + F_3P = G_1P + NDP + E_3 + F_3 + 2\ P \tag{14}$$

Again, the AND-connection leads to only one EFM, as expected. The EFM consists of all reactions in the model, and its overall reaction looks as follows:

$$E_1P + F_1P + G_3 + 7\ NTP = E_1 + F_1 + G_3P + 7\ NDP + 8\ P \tag{15}$$

3.2. Double-phosphorylation cascades

A further motif in signal transduction networks are double-phosphorylation cascades. In such cascades, an enzyme E_{i+1} needs to be phosporylated twice to become active.

$$E_{i+1} + NTP = E_{i+1}P + NDP \tag{16a}$$

$$E_{i+1}P + NTP = E_{i+1}PP + NDP \tag{16b}$$

These two phosphorylation steps are often catalyzed by the same preceding enzyme, just as the two subsequent dephosphorylation steps are usually catalyzed by the same phosphatase. Without any kinetic

Table 2. Metatool Input for the OR-Connected Converging Cascade Shown in Figure 2

```
-ENZREV

-ENZIRREV
R1 R2 R3 R4 R5 R6 R7 R8

-METINT
E2 E2P E3 E3P F2 F2P F3 F3P G1 G1P G2 G2P

-METEXT
NDP NTP E1 E1P F1 F1P G3 G3P P

-CAT
R1 : 1 NTP + 1 E2 + 1 E1P = 1 E2P + 1 NDP + 1 E1 + 1 P
R2 : 1 NTP + 1 E3 + 1 E2P = 1 E3P + 1 NDP + 1 E2 + 1 P
R3 : 1 NTP + 1 F2 + 1 F1P = 1 F2P + 1 NDP + 1 F1 + 1 P
R4 : 1 NTP + 1 F3 + 1 F2P = 1 F3P + 1 NDP + 1 F2 + 1 P
R5 : 1 NTP + 1 G1 + 1 E3P = 1 G1P + 1 NDP + 1 E3 + 1 P
R6 : 1 NTP + 1 G1 + 1 F3P = 1 G1P + 1 NDP + 1 F3 + 1 P
R7 : 1 NTP + 1 G2 + 1 G1P = 1 G2P + 1 NDP + 1 G1 + 1 P
R8 : 1 NTP + 1 G3 + 1 G2P = 1 G3P + 1 NDP + 1 G2 + 1 P
```

constants, the exact chronology of phosphorylation and dephosphorylation, and thus the amplification of the signal along the cascade, cannot be determined. But again a minimal precondition to transduce a signal can be identified. That is, we can assume that an enzyme E_i is double-phosporylated and stays in this active state until the next enzyme E_{i+1} is phosporylated twice. Therefore, we can merge the two phosphorylation steps and set up eq. (17) analogously to eq. (5).

$$E_iPP + n\ E_{i+1} + 2n\ NTP = E_i + 2\ P + n\ E_{i+1}PP + 2n\ NDP \quad (17)$$

Again, we have combined the phosphorylation of E_{i+1} and the subsequent dephosphorylation of E_i into one reaction. In the minimum case of no signal amplification, $n = 1$, we get:

$$E_iPP + E_{i+1} + 2\ NTP = E_i + 2\ P + E_{i+1}PP + 2\ NDP \quad (18)$$

This reaction equation applies to all levels i of the cascade.

The double-phosphorylation cascade and its simplified form with merged phosphorylation steps are shown in Figures 3 and 4, respectively. The latter scheme gives rise to one EFM. The corresponding overall reaction reads as follows:

$$E_1PP + E_4 + 6\ NTP = E_1 + E_4PP + 6\ NDP + 6\ P \quad (19)$$

The complete Metatool input is given in Table 3.

Sometimes a single-phosphorylated kinase $E_{i+1}P$ shows a different substrate specificity than the double-phosphorylated $E_{i+1}PP$. In such cases, the $E_{i+1}P$ can catalyze the phosphorylation of a kinase F_1 while $E_{i+1}PP$ phosphorylates G_1, and thus a diverging double-phosphorylation cascade with a branching point occurs. But again we do not know any kinetics and thus can model neither the amplification of the signal along the cascade nor the exact interplay of all phosphorylation and dephosphorylation steps of E_{i+1}. Nevertheless, the following two minimal preconditions must be valid: To transduce the signal along the two possible routes of the diverging cascade, it is necessary that F_1 is double-phosphorylated by $E_{i+1}P$ and G_1 by $E_{i+1}PP$. So we reduce the branching point to the following four reactions:

$$E_iPP + n\ E_{i+1} + n\ NTP = E_i + 2\ P + n\ E_{i+1}P + n\ NDP \quad (20a)$$

$$E_iPP + n\ E_{i+1} + 2n\ NTP = E_i + 2\ P + n\ E_{i+1}PP + 2n\ NDP \quad (20b)$$

$$E_{i+1}P + n\ F_1 + 2n\ NTP = E_{i+1} + P + n\ F_1PP + 2n\ NDP \quad (20c)$$

$$E_{i+1}PP + n\ G_1 + 2n\ NTP = E_{i+1} + 2\ P + n\ G_1PP + 2n\ NDP \quad (20d)$$

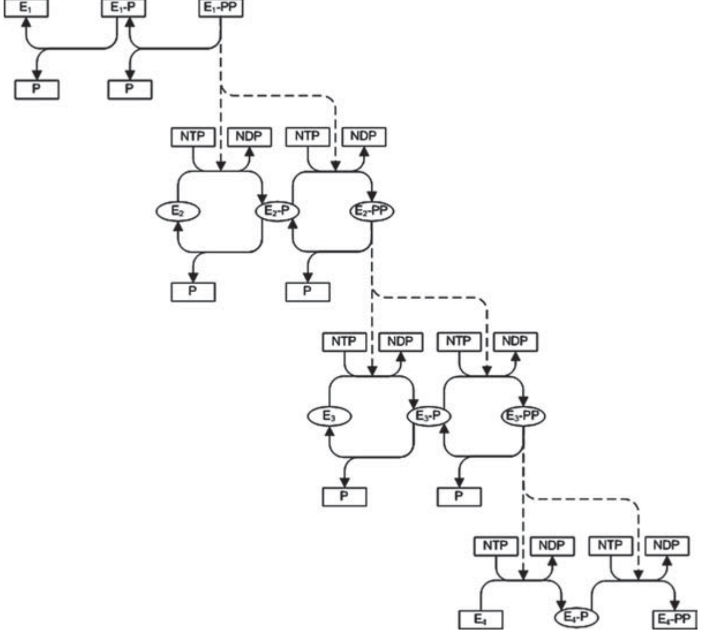

FIG. 3. Double-phosphorylation cascade, comprising the kinases E_1–E_4. Rectangles, external metabolites; ellipses, internal metabolites.

Again, we have combined the phosphorylation of E_{i+1} and the subsequent dephosphorylation of E_i. In the minimum case $n = 1$, we get:

$$E_iPP + E_{i+1} + NTP = E_i + 2\ P + E_{i+1}P + NDP \quad (21a)$$

$$E_iPP + E_{i+1} + 2\ NTP = E_i + 2\ P + E_{i+1}PP + 2\ NDP \quad (21b)$$

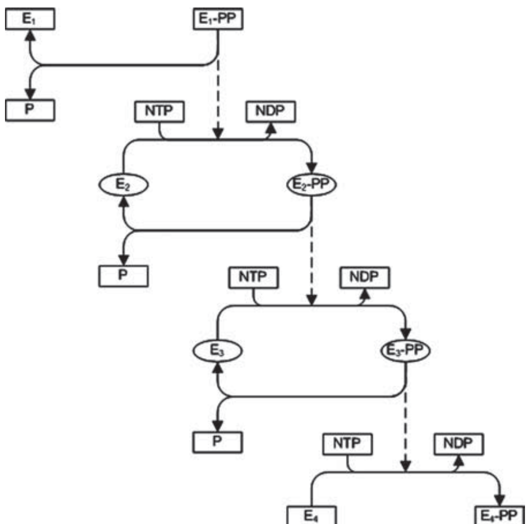

FIG. 4. Simplified double-phosphorylation cascade, comprising the kinases E_1–E_4. Rectangles, external metabolites; ellipses, internal metabolites.

$$E_{i+1}P + F_1 + 2\ NTP = E_{i+1} + P + F_1PP + 2\ NDP \tag{21c}$$

$$E_{i+1}PP + G_1 + 2\ NTP = E_{i+1} + 2\ P + G_1PP + 2\ NDP \tag{21d}$$

For the other reaction equations of the cascade, again equations analogous to eq. (18) can be used.

The diverging double-phosphorylation cascade and its simplified form with merged phosphorylation steps are shown in Figures 5 and 6, respectively. The latter scheme gives rise to two EFMs, consisting of the reactions R_1, R_2, R_4, R_6, and R_7, and of R_1, R_3, R_5, R_8, and R_9. The corresponding overall reactions are:

TABLE 3. METATOOL INPUT FOR THE SIMPLIFIED DOUBLE-PHOSPHORYLATION CASCADE SHOWN IN FIGURE 4

```
-ENZREV

-ENZIRREV
R1 R2 R3

-METINT
E2 E2PP E3 E3PP

-METEXT
NDP NTP E1 E1PP E4 E4PP P

-CAT
R1 : 2 NTP + 1 E2 + 1 E1PP = 1 E2PP + 2 NDP + 1 E1 + 2 P
R2 : 2 NTP + 1 E3 + 1 E2PP = 1 E3PP + 2 NDP + 1 E2 + 2 P
R3 : 2 NTP + 1 E4 + 1 E3PP = 1 E4PP + 2 NDP + 1 E3 + 2 P
```

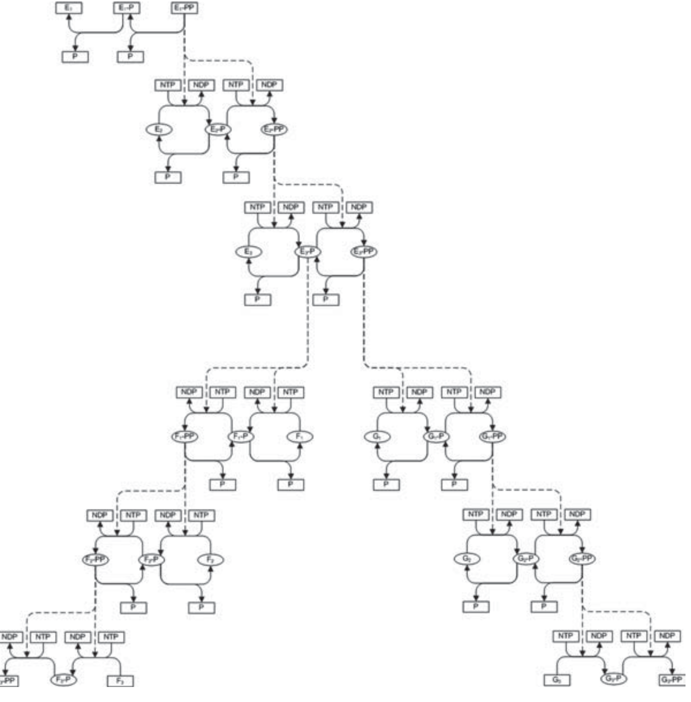

FIG. 5. Diverging double-phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3, and G_1–G_3. Rectangles, external metabolites; ellipses, internal metabolites.

$$E_1PP + F_3 + 9\ NTP = E_1 + F_3PP + 9\ NDP + 9\ P \quad (22a)$$

$$E_1PP + G_3 + 10\ NTP = E_1 + G_3PP + 10\ NDP + 10\ P \quad (22b)$$

The complete Metatool input is given in Table 4.

4. EXAMPLE FROM INSULIN SIGNALING

As a real example for a diverging cascade, we modeled part of the insulin signaling pathway in humans based on data from the Transpath database (Krull et al., 2006). The model (Metatool input given in Table 5) starts with a binding reaction between insulin and its receptor:

$$\text{Insulin} + \text{InsR} = \text{IIR} \quad (23)$$

where IIR denotes the insulin-insulin-receptor complex. This complex phosphorylates itself and becomes active. After a binding reaction with Shc, a subsequent phosphorylation and a further complexation with Grb-2 and Sos, the resulting enzyme complex is able to catalyze the GDP-GTP-exchange reaction that activates Ras. Formally, the exchange of GDP and GTP can be modeled in the same way as a

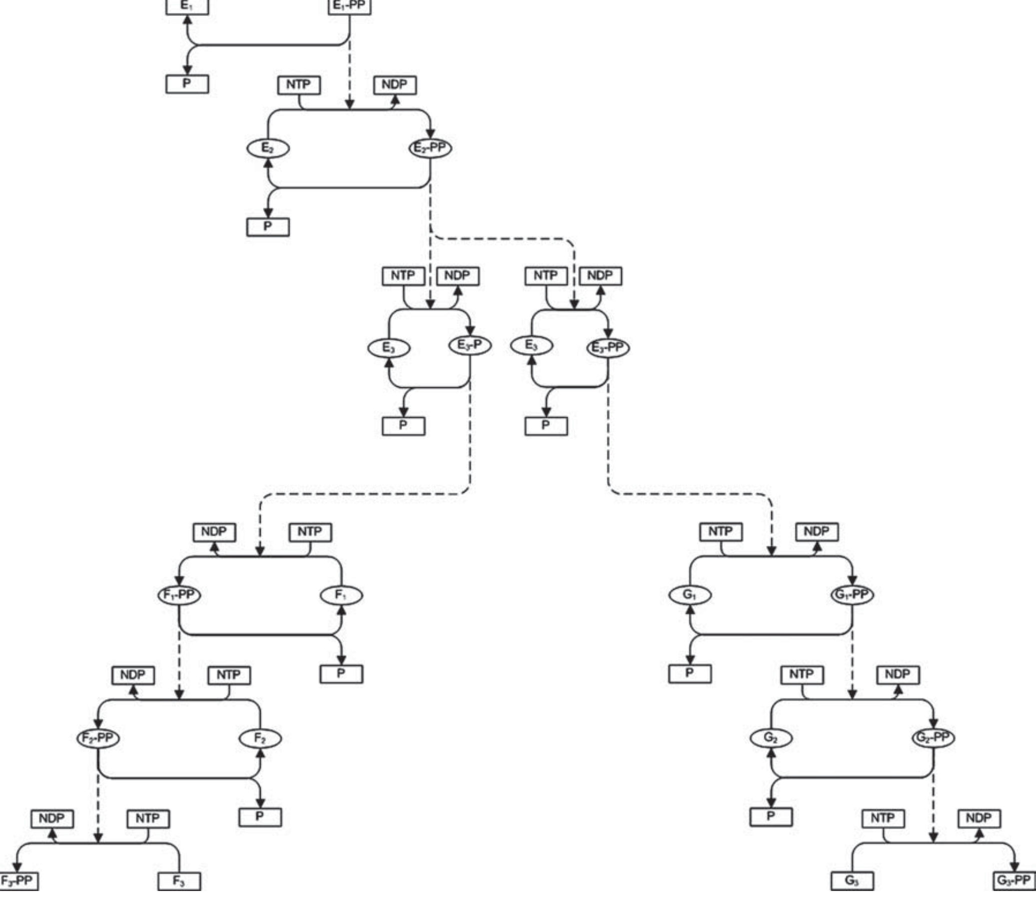

FIG. 6. Simplified diverging double-phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3, and G_1–G_3. Rectangles, external metabolites; ellipses, internal metabolites.

TABLE 4. METATOOL INPUT FOR THE SIMPLIFIED DIVERGING DOUBLE-PHOSPHORYLATION CASCADE SHOWN IN FIGURE 6

```
-ENZREV

-ENZIRREV
R1 R2 R3 R4 R5 R6 R7 R8 R9

-METINT
E2 E2PP E3 E3P E3PP F1 F1PP F2 F2PP G1 G1PP G2 G2PP

-METEXT
NDP NTP E1 E1PP F3 F3PP G3 G3PP P

-CAT
R1 : 2 NTP + 1 E2 + 1 E1PP = 1 E2PP + 2 NDP + 1 E1 + 2 P
R2 : 1 NTP + 1 E3 + 1 E2PP = 1 E3P  + 1 NDP + 1 E2 + 2 P
R3 : 2 NTP + 1 E3 + 1 E2PP = 1 E3PP + 2 NDP + 1 E2 + 2 P
R4 : 2 NTP + 1 F1 + 1 E3P  = 1 F1PP + 2 NDP + 1 E3 + 1 P
R5 : 2 NTP + 1 G1 + 1 E3PP = 1 G1PP + 2 NDP + 1 E3 + 2 P
R6 : 2 NTP + 1 F2 + 1 F1PP = 1 F2PP + 2 NDP + 1 F1 + 2 P
R7 : 2 NTP + 1 F3 + 1 F2PP = 1 F3PP + 2 NDP + 1 F2 + 2 P
R8 : 2 NTP + 1 G2 + 1 G1PP = 1 G2PP + 2 NDP + 1 G1 + 2 P
R9 : 2 NTP + 1 G3 + 1 G2PP = 1 G3PP + 2 NDP + 1 G2 + 2 P
```

TABLE 5. METATOOL INPUT FOR THE MODELED PART OF THE INSULIN PATHWAY SHOWN IN FIGURE 7

```
-ENZREV
R01 R03 R05 R06 R09

-ENZIRREV
R02 R04 R07 R08 R10 R11 R12 R13 R14 R15

-METINT
ERK ERK(P) Grb–2 Grb–2_Sos IIR IIR(PY) IIR(PY)_Shc IIR(PY)_Shc(PY) IIR(PY)_Shc(PY)_Grb−2_Sos
  InsR Insulin MEK MEK(P) RSK RSK(P) Raf Raf(P) Ras_GDP Ras_GTP Ras_GTP_Raf(P) Shc Sos

-METEXT
ADP ATP CREB CREB(PS) GDP GTP NDP NTP P PFKFB–2 PFKFB−2(PS) S6 S6(P)

-CAT
R01 : 1 Insulin + 1 InsR = 1 IIR
R02 : 1 IIR + 1 ATP = 1 IIR(PY) + 1 ADP
R03 : 1 IIR(PY) + 1 Shc = 1 IIR(PY)_Shc
R04 : 1 IIR(PY)_Shc + 1 NTP = 1 IIR(PY)_Shc(PY) + 1 NDP
R05 : 1 Grb–2 + 1 Sos = Grb−2_Sos
R06 : 1 IIR(PY)_Shc(PY) + 1 Grb–2_Sos = IIR(PY)_Shc(PY)_Grb–2_Sos
R07 : 1 IIR(PY)_Shc(PY)_Grb–2_Sos + 1 Ras_GDP + 1 GTP = 1 Ras_GTP + 1 GDP + 1 Insulin + 1 InsR +
      1 Shc + 1 Grb–2 + 1 Sos + 2 P
R08 : 1 Raf + 1 ATP = 1 Raf(P) + 1 ADP
R09 : 1 Ras_GTP + 1 Raf(P) = Ras_GTP_Raf(P)
R10 : 1 Ras_GTP_Raf(P) + 1 MEK + 1 NTP = 1 MEK(P) + 1 NDP + 1 Ras_GDP + 1 Raf + 2 P
R11 : 1 ERK + 1 NTP + 1 MEK(P) = 1 ERK(P) + 1 NDP + 1 MEK + 1 P
R12 : 1 RSK + 1 NTP + 1 ERK(P) = 1 RSK(P) + 1 NDP + 1 ERK + 1 P
R13 : 1 PFKFB−2 + 1 ATP + 1 RSK(P) = 1 PFKFB−2(PS) + 1 ADP + 1 RSK + 1 P
R14 : 1 CREB + 1 ATP + 1 RSK(P) = 1 CREB(PS) + 1 ADP + 1 RSK + 1 P
R15 : 1 S6 + 1 ATP + 1 RSK(P) = 1 S6(P) + 1 ADP + 1 RSK + 1 P
```

P, phosphorylation in general; PS, phosphorylation at a serine; PY, phosphorylation at a tyrosine.

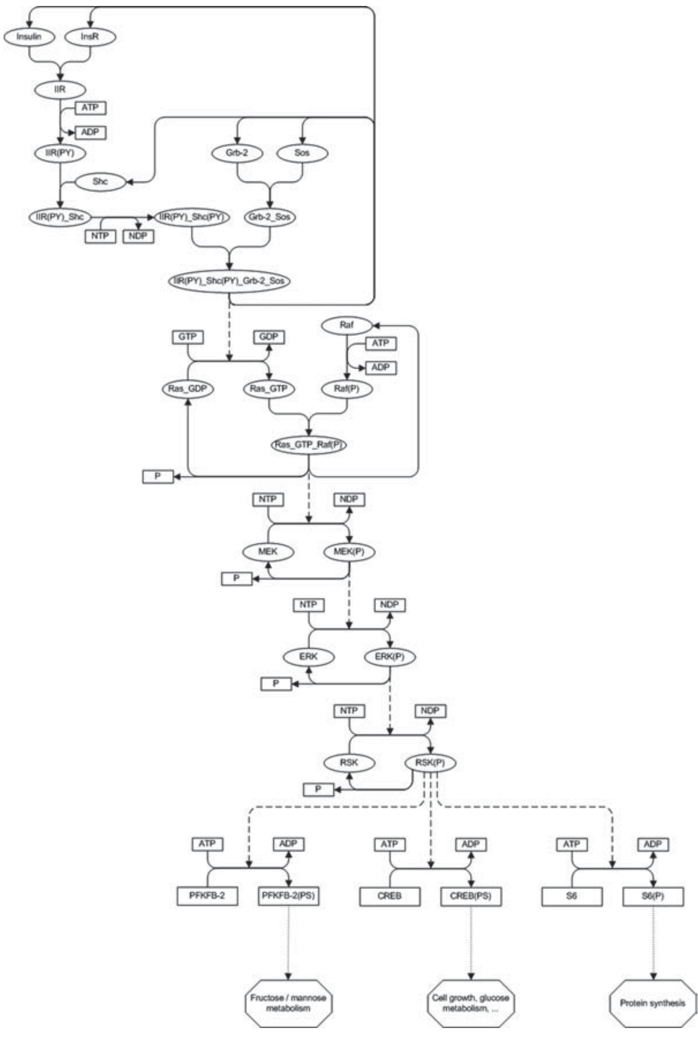

FIG. 7. Diverging cascade taken from the insulin signaling pathway. The complete description of the components involved is given in the Appendix. Rectangles, external metabolites; ellipses, internal metabolites. Octagons represent those parts of the cell that are finally affected by the considered signal transduction network.

phosphorylation of GDP by GTP. The kinase cascade continues via MEK, ERK, and RSK. The branching point is given by the kinase RSK, which catalyzes (in its phosphorylated form) the phosphorylation (and hence activation) of the enzyme PFKFB-2, the transcription factor CREB and the ribosomal protein S6. PFKFB-2 is a bifunctional enzyme, known as 6-phosphofructo-2-kinase (EC 2.7.1.105) and fructose-2,6-bisphosphate 2-phosphatase (EC 3.1.3.46), both being important in hexose metabolism. CREB and S6 affect glycolysis and protein synthesis, respectively. The corresponding reactions are labeled R_{13}–R_{15} in Table 5. The modeled cascade is shown in Figure 7.

This system gives rise to three EFMs. They have reactions R_{01}–R_{12} in common and differ only in the last reaction (R_{13}, R_{14}, R_{15}). The corresponding overall reactions read as follows:

$$3\ \mathrm{ATP} + \mathrm{GTP} + 4\ \mathrm{NTP} + \mathrm{PFKFB{-}2} = 3\ \mathrm{ADP} + \mathrm{GDP} + 4\ \mathrm{NDP} + 7\ \mathrm{P} + \mathrm{PFKFB{-}2(PS)} \tag{24a}$$

$$3\ \mathrm{ATP} + \mathrm{CREB} + \mathrm{GTP} + 4\ \mathrm{NTP} = 3\ \mathrm{ADP} + \mathrm{CREB(PS)} + \mathrm{GDP} + 4\ \mathrm{NDP} + 7\ \mathrm{P} \tag{24b}$$

$$3\ \mathrm{ATP} + \mathrm{GTP} + 4\ \mathrm{NTP} + \mathrm{S6} = 3\ \mathrm{ADP} + \mathrm{GDP} + 4\ \mathrm{NDP} + 7\ \mathrm{P} + \mathrm{S6(P)} \tag{24c}$$

These three EFMs can be interpreted in biological terms as follows. Insulin activates fructose/mannose metabolism, glycolysis as well as protein synthesis. Although these three effects usually occur simultaneously, each EFM could operate alone when the other two are non-functional, for example, when some component proteins are knocked out. For this system, the elementary signaling routes are quite obvious. We have taken this example as a proof of concept. In larger networks, the signaling routes can no longer be detected by inspection.

5. DISCUSSION

Here, we have presented an approach for detecting elementary signaling routes in enzyme cascades (e.g., phosphorylation cascades). To this end, we adapted elementary flux modes (EFM) analysis, which had been established earlier for detecting pathways in metabolism (Schuster et al., 1999, 2000). A schematic application of the concept of EFMs to enzyme cascades would lead to the trivial result that the EFMs reflect particular enzyme cycles (e.g., phosphorylation-dephosphorylation cycles) rather than the routes of information transfer. In order to describe the latter in a suitable way, we have started from the reasoning that signaling usually implies signal amplification or at least a constant signal strength, but never a diminuition. In the limit case of constant signal strength, each active enzyme molecule should activate exactly one enzyme molecule at the next level of the cascade. By this coupling, we can write reaction equations that lead to EFMs representing routes of information transfer. The same routes are relevant when signal amplification occurs. By this reasoning, the application of elementary flux modes analysis and related methods such as extreme pathway analysis (Schilling et al., 2000) and minimal T-invariants (Starke, 1990) to intracellular signaling systems has been put on a firm theoretical basis. Thus, EFMs can be calculated, for example, by the program Metatool (von Kamp and Schuster, 2006) also for enzyme cascades. Earlier, these analyses have been applied in a formal way to such systems without a theoretical justification (Xiong et al., 2004; Heiner et al., 2004; Sackmann et al., 2006).

As a proof of concept, we have applied the presented method to part of the insulin signaling network. The three resulting EFMs can be interpreted in biochemical terms. The first EFM leads to a serine phosphorylated and thus active 6-phosphofructo-2-kinase/fructose 2,6-bisphosphatase (EC 3.1.3.46/EC 2.7.1.105). This bifunctional enzyme is part of the fructose and mannose metabolism, where it catalyzes the reaction from β-D-fructose 6-phosphate to β-D-fructose 2,6-bisphosphate, and back. The output of the second EFM is the activated transcription factor CREB(PS), which regulates (together with the transcription factor forkhead box O1A) the expression of the insulin-like growth factor binding protein 1. That protein is neccessary for binding the insulin-like growth factor 1, which in turn regulates cell growth, glucose metabolism, and several other processes. The third EFM results in the phosphorylated ribosomal protein S6. Hence, the further outcome of this signaling route is activation of protein synthesis. For all of these functions, see the Transpath database (Krull et al., 2006).

Of course, the entire insulin signaling network has many more functions, such as activation of glycogenesis and enhancing glucose uptake in muscle cells and adipocytes by increasing the number of GLUT4 transporters in their membranes.

The enzyme cascades analyzed here can be considered as relay races with changing batons because it might appear as if the phosphates were transferred along the cascades, but they are replaced at each level. This situation differs from group transfer pathways such as the phophotransferase system (PTS) for which the analogy to a relay race is more appropriate because the "baton" is not replaced. Papin and Palsson (2004) have studied the latter type of systems.

The enzyme Ras, which we have included in our model, belongs to the G-proteins and in particular to the family of small GTPases. Such enzymes are activated by replacing the GDP they are complexed with in their inactive form, by GTP. The substitution is catalyzed by so-called guanine nucleotide exchange factors (GEFs) (Krauss, 2003). In our case, this is IIR(PY)_Shc(PY)_Grb-2_Sos, the complex of insulin, insulin receptor, Src homologous and collagen protein, growth-factor receptor-binding protein 2, and the enzyme Son of Sevenless. The inactivation of these activated G-proteins is triggered by their intrinsic ability to hydrolyze GTP to GDP and phosphate (Gomperts et al., 2002). From the point of view of the biochemical structure of the active form of the G-protein, the replacement of GDP by GTP is equivalent to a phosphorylation. Therefore, our formalism is also applicable to this class of signaling proteins.

Our analysis differs from the approach proposed by Alon (2007) in that we determine routes going through the entire signaling network, starting from an initial signal and leading to some cellular response. In contrast,

Alon (2007) extracts modular (non-overlapping) network motifs, which are smaller than the signaling routes determined here, which can have subroutes in common.

It is promising to apply robustness analysis (Wilhelm et al., 2004; Behre et al., 2008) and the concept of minimal cut sets (Klamt and Gilles, 2004; Klamt, 2006) proposed earlier for metabolic networks to enzyme cascades. Since these methods use EFMs, their application is straightforward. Also the concept of enzyme subsets (Pfeiffer et al., 1999) and the refined concepts in flux coupling analysis (Burgard et al., 2004) can be applied. On Boolean networks, information transfer has been analyzed, and minimal intervention sets have been defined without explicitly considering mass balance constraints (Klamt et al., 2006). It would be interesting to investigate the interrelations between these approaches.

6. APPENDIX

Components involved in the modeled part of the human insulin pathway are described in the following table. P denotes phosphorylation in general; PS, phosphorylation at a serine; and PY, phosphorylation at a tyrosine.

Abbreviated component name	*Description*
ADP	Adenosine 5′-diphosphate
ATP	Adenosine 5′-triphosphate
CREB	cAMP-responsive element-binding protein
CREB(PS)	cAMP-responsive element-binding protein (serine phosphorylated)
ERK	Extracellular signal-regulated kinase 1
ERK(P)	Extracellular signal-regulated kinase 1 (phosphorylated)
GDP	Guanosine 5′-diphosphate
Grb-2	Growth-factor receptor-binding protein 2
Grb-2_Sos	Complex of Grb-2 and Sos
GTP	Guanosine 5′-triphosphate
IIR	Insulin insulin receptor complex
IIR(PY)	Insulin insulin receptor complex (tyrosine phosphorylated)
IIR(PY)_Shc	Complex of IIR(PY) and Shc
IIR(PY)_Shc(PY)	Complex of IIR(PY) and Shc (tyrosine phosphorylated)
IIR(PY)_Shc(PY)_Grb-2_Sos	Complex of IIR(PY)_Shc(PY) and Grb-2_Sos
InsR	Insulin receptor
Insulin	Insulin
MEK	ERK-kinase (mitogen-activated protein kinase)
MEK(P)	ERK-kinase (phosphorylated)
NDP	Nucleoside diphosphate
NTP	Nucleoside triphosphate
P	Orthophosphate
PFKFB-2	6-Phosphofructo-2-kinase/fructose 2,6-bisphosphatase (EC 3.1.3.46 / EC 2.7.1.105)
PFKFB-2(PS)	PFKFB-2 (serine phosphorylated)
Raf	Raf-1, serine/threonine-specific kinase (EC 2.7.11.1) Raf-1 is a MAP kinase kinase kinase (MAP3K).
Raf(P)	Raf (phosphorylated)
Ras_GDP	GTPase from the Ras family (inactive form)
Ras_GTP	GTPase from the Ras family (active form)
Ras_GTP_Raf(P)	Complex of Ras_GTP and Raf(P)
RSK	Ribosomal S6 kinase
RSK(P)	Ribosomal S6 kinase (phosphorylated)
S6	Ribosomal protein S6
S6(P)	Ribosomal protein S6 (phosphorylated)
Shc	Src homologous and collagen protein (adaptor protein)
Sos	Son of Sevenless (GTPase-controlling signal molecule)

ACKNOWLEDGMENTS

We thank Steffen Klamt, Ina Koch, and Regina Samaga for stimulating discussions. Financial support by the BMBF (German Ministry for Education and Research) to J. Behre (HepatoSys Program and Jena Center for Bioinformatics) is gratefully acknowledged.

DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

Alon, U. 2007. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* Chapman & Hall/CRC, Boca Raton, FL.

Arga, K.Y., Önsan, Z.I., Kirdar, B., et al. 2007. Understanding signaling in yeast: insights from network analysis. *Biotechnol. Bioeng.* 97, 1246–1258.

Behre, J., Wilhelm, T., von Kamp, A., et al. 2008. Structural robustness of metabolic networks with respect to multiple knockouts. *J. Theor. Biol.* 252, 433–441.

Burgard, A.P., Nikolaev, E.V., Schilling, C.H., et al. 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res.* 14, 301–312.

Clarke, B.L. 1981. Complete set of steady states for the general stoichiometric dynamical system. *J. Chem. Phys.* 75, 4970–4979.

Gomperts, B.D., Kramer, I.M., and Tatham, P.E.R. 2002. *Signal Transduction.* Academic Press, San Diego.

Heiner, M., Koch, I., and Will, J. 2004. Model validation of biological pathways using Petri nets—demonstrated for apoptosis. *Biosystems* 75, 15–28.

Heinrich, R., and Schuster, S. 1996. *The Regulation of Cellular Systems.* Chapman & Hall, New York.

Klamt, S. 2006. Generalized concept of minimal cut sets in biochemical networks. *Biosystems* 83, 233–247.

Klamt, S., and Gilles, E.D. 2004. Minimal cut sets in biochemical reaction networks. *Bioinformatics* 20, 226–234.

Klamt, S., Saez-Rodriguez, J., Lindquist, J.A., et al. 2006. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinform.* 7, 56.

Krauss, G. 2003. *Biochemistry of Signal Transduction and Regulation.* 3rd ed. Wiley-VCH, Weinheim, Germany.

Krull, M., Pistor, S., Voss, N., et al. 2006. TRANSPATH®: an Information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Res.* 34, D546–D551.

Papin, J.A., and Palsson, B.Ø. 2004. Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *J. Theor. Biol.* 227, 283–297.

Papin, J.A., Stelling, J., Price, N.D., et al. 2004. Comparison of network-based pathway analysis methods. *Trends Biotechnol.* 22, 400–405.

Pfeiffer, T., Sanchez-Valdenebro, I., Nuño, J.C., et al. 1999. METATOOL: for studying metabolic networks. *Bioinformatics* 15, 251–257.

Sackmann, A., Heiner, M., and Koch, I. 2006. Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinform.* 7, 482.

Sackmann, A., Formanowicz, D., Formanowicz, P., et al. 2007. An analysis of the Petri net based model of the human body iron homeostasis process. *Comput. Biol. Chem.* 31, 1–10.

Savageau, M.A. 1976. *Biochemical Systems Analysis*. Addison-Wesley, Reading, MA.

Schilling, C.H., Letscher, D., and Palsson, B.Ø. 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229–248.

Schuster, S., and Hilgetag, C. 1994. On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.* 2, 165–182.

Schuster, S., Dandekar, T., and Fell, D.A. 1999. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends Biotechnol.* 17, 53–60.

Schuster, S., Fell, D.A., and Dandekar, T. 2000. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18, 326–332.

Schuster, S., Hilgetag, C., Woods, J.H., et al. 2002. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* 45, 153–181.

Schwartz, J.M., Gaugain, C., Nacher, J.C., et al. 2007. Observing metabolic functions at the genome scale. *Genome Biol.* 8, R123.1–R123.17.

Schwender, J., Goffman, F., Ohlrogge, J.B., et al. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432, 779–782.

Segel, I.H. 1993. *Enzyme Kinetics.* Wiley, New York.

Starke, P.H. 1990. *Analyse von Petri-Netz-Modellen.* Teubner, Stuttgart.

Steffen, M., Petti, A., Aach, J., et al. 2002. Automated modelling of signal transduction networks. *BMC Bioinform.* 3, 34.

Stelling, J., Klamt, S., Bettenbrock, K., et al. 2002. Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, 190–193.

von Kamp, A., and Schuster, S. 2006. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics* 22, 1930–1931.

Wilhelm, T., Behre, J., and Schuster, S. 2004. Analysis of structural robustness of metabolic networks. *IEE Proc. Syst. Biol.* 1, 114–120.

Xiong, M., Zhao, J., and Xiong, H. 2004. Network-based regulatory pathways analysis. *Bioinformatics* 20, 2056–2066.

Zevedei-Oancea, I., and Schuster, S. 2003. Topological analysis of metabolic networks based on Petri net theory. *In Silico Biol.* 3, 323–345.

Zevedei-Oancea, I., and Schuster, S. 2005. A theoretical framework for detecting signal transfer routes in signalling networks. *Comput. Chem. Eng.* 29, 597–617.

Address reprint requests to:
Jörn Behre
Faculty of Biology and Pharmaceutics
Section of Bioinformatics
Friedrich Schiller University Jena
Ernst-Abbe-Platz 2
D-07743 Jena, Germany

E-mail: joern.behre@uni-jena.de

Chapter 5

General Discussion

In this PhD thesis, I have presented a collection of three articles focusing on the calculation of the structural analysis of metabolic and signal transduction networks, and on the comparison of different networks and their topologies, with the help of new measures for structural robustness that we introduced.

Structural robustness based on single knockouts

In our first paper (Wilhelm *et al.*, 2004, Chapter 2), we showed that the approach of Stelling *et al.* (2002), to take the number of EFMs as a measure of robustness, is not sufficient. Fig. 5.1, taken from Wilhelm *et al.* (2004), shows that networks can have the same number of EFMs but a different robustness. By considering exclusively the number of EFMs, it remains unclear what happens to the topology of the considered network if an enzyme is knocked out. Also the approach of Çakir *et al.* (2004a,b) to compare the biomass yields of wild type and mutant strains and to take the biomass yield as a measure of fault-tolerance of these cells, is of limited use, since not all metabolic networks, the robustness of which shall be analysed, produce biomass.

We proposed three new measures of structural robustness R_1, R_2, and R_3 (see Wilhelm *et al.*, 2004, Chapter 2), based on a comparison of numbers of EFMs in the states before and after knockouts. We introduced three different variants to cope also with subnetworks. Whereas for the general definition, R_1, all EFMs of the entire network are considered, the measures R_2 and R_3 are based on subsets of EFMs leading to products P_k that are of special

interest. In the following, these products can be considered separately to calculate so-called *essential product robustnesses* $R_1^{(P_k)}$. In the case of all of these products being essential to an organism, the resulting robustness R_2 needs to be calculated as the minimum of all $R_1^{(P_k)}$, otherwise (for R_3) the usual average is sufficient.

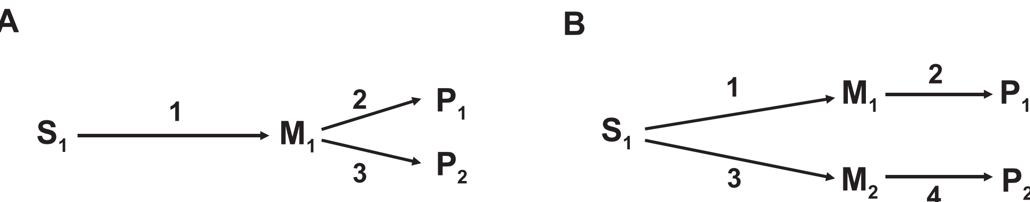

Figure 5.1. Example networks, taken from Wilhelm *et al.* (2004), with the same number of EFMs but a different robustness. Network **A** is less robust than **B**, since a single knockout of reaction 1, forming a kind of Achilles heel, can cut off any EFM in **A**, while at least two knockouts are needed in **B** to remove all EFMs. However, in order to be more robust, an additional reaction is required in network **B** indicating the cost of robustness. M_i are internal metabolites, S_k substrates and P_m products.

Although the measures we introduced are normalised quantities between 0 and 1 and thus independent of the network size, we compared in our first study in Chapter 2 different networks of *Escherichia coli* and human erythrocyte with the same number of essential metabolites and almost the same number of EFMs, to be on the safe side and exclude any effect of these factors. The values in Table 1 in Wilhelm *et al.* (2004) show just small deviations within the set of R_1 robustness values for the different *Escherichia coli* networks (0.51 to 0.55), whereas the corresponding value for human erythrocyte is markedly lower (0.38). We also analysed another erythrocyte model published by Joshi and Palsson (1989) that comprises 21 EFMs. With hypoxanthine excretion and sodium/potassium transport as essential functions this model gives rise to the following robustness values: $R_1 = 0.44$, $R_2 = 0.20$, and $R_3 = 0.21$. Their difference to the values for *Escherichia coli* is even more pronounced. These differences in robustness are in accordance with common biochemical knowledge saying that erythrocytes have a much simpler and, hence, less robust metabolism than *Escherichia coli*. The latter needs to be able to adapt to different conditions such as the human intestine or varying environments outside the human body, whereas human erythrocytes live under much more homeostatic conditions in the blood. The

values for R_2 and R_3 are always smaller than that for R_1, but with the same robustness trend. This effect is caused by a basic phenomenon in structural analysis that always emerges when cut-outs of larger networks are considered. For the calculation of R_2 and R_3, subsets of EFMs and thus subnetworks are used. They are always pruned and thus less interconnected compared with the entire network, and consequently give rise to lower structural robustness. On the other hand, the observation that values of R_2 are always smaller than those of R_3, is plausible, since R_2 and R_3 are defined as the minimum and average values of the same set of quantities, respectively. So, the inequality $R_2 \leq R_3$ must hold.

In their study, Ebenhöh and Heinrich (2003) distinguish between strong robustness and weak robustness. According to their definition, a metabolic network is strongly robust against a certain mutation if it still can produce the same products. It is weakly robust if it can still produce at least one product. Within our framework, a strongly robust network would have a robustness measure R_2 larger than zero (assumed that all products are considered to be essential), while for a weakly robust network, R_2 would be equal to zero.

The question arises whether it is actually allowed to compare different parts from different metabolisms in different organisms. Does it make sense or is it a comparison between apples and oranges? Since we are not comparing the different networks with respect to their functionalities but concerning their different topologies, such comparisons make sense. The two small networks in Fig. 5.2 have obviously the same topology and both have a structural robustness of $R_1 = 0.5$. But their functionality is totally different, since network **A** converts the substrate S_1 into the products P_1 and P_2, whereas in network **B**, the products Z_1 and Z_2 are produced from substrate X_1.

Experimental observations show that most enzyme deficiencies in erythrocytes concerning glycolysis (e.g. hexokinase, pyruvate kinase, or glucose-6-phosphate dehydrogenase) lead to severe diseases (Scriver *et al.*, 1995), whereas the majority of single-gene deletions of enzymes in *Escherichia coli* does not imply inviability (Ingraham and Neidhardt, 1987). A further analysis of the 667 EFMs calculated for the erythrocyte model reveals that 400 EFMs use hexokinase, 528 EFMs pyruvate kinase, and 428 EFMs glucose-6-phosphate dehydrogenase. This is a quite obvious explanation for the lethal-

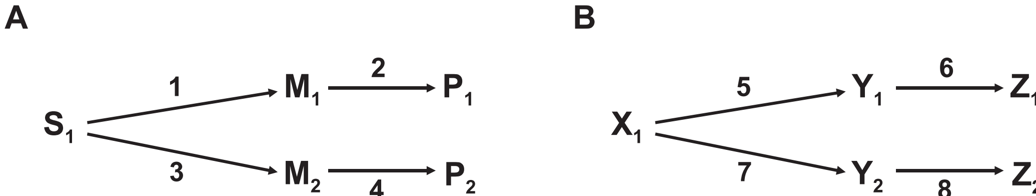

Figure 5.2. Two example networks with different functionalities, but having the same topologies and thus the same robustness. Network **A** converts the substrate S_1 into the products P_1 and P_2, whereas in network **B**, the products Z_1 and Z_2 are produced from substrate X_1. M_i, Y_i are internal metabolites, S_k, X_k substrates, and P_m, Z_m products.

ity of these knockouts. In contrast, *Escherichia coli* is able to use different nutrients (Ingraham and Neidhardt, 1987).

The relation between the *fragility coefficient* introduced by Klamt and Gilles (2004) within the framework of MCSs and our concept of structural robustness is not straightforward, since converse approaches are used to assess the network fragility and robustness, respectively. From our robustness measures also fragility measures can be derived as $1-R_1$, $1-R_2$, and $1-R_3$, respectively. In contrast, the fragility coefficient F_i introduced by Klamt and Gilles (2004) is defined as the reciprocal of the average size of all MCSs in which a reaction (enzyme) E_i participates. The related network (overall) fragility coefficient F is defined as the average of all fragility coefficients F_i over all reactions E_i. Nevertheless, since the fragility coefficient F_i is based on MCSs that are calculated for a certain objective function, both concepts are at best similar if the objective function (e.g. reaction E_x) coincides with the chosen subset of EFMs the robustness calculations are based on (e.g. all EFMs in which reaction E_x participates). Moreover, since the MCSs F_i is based on single and multiple knockouts, they are not quite comparable with the systematically processed single, double, and multiple knockouts in our approach. Nevertheless, to assess the fragility of a system concerning a certain objective reaction (e.g. reaction E_x) and a certain single reaction of interest (e.g. reaction E_y), the fragility coefficient F_y is well suited. This is also possible within our robustness framework by taking the subset of EFMs, in which reaction E_x participates, and to determine $1-z^{(y)}/z$, the percentage of EFMs that drop out when reaction E_y is blocked. $z^{(y)}/z$ is computed in any case during the calculation of R_1, R_2, and R_3, respectively.

To analyse the fraction $z^{(i)}/z$ of remaining EFMs after knockout of reaction E_i separately for all blockable reactions in the system under study is a good way to determine bottlenecks (or Achilles heels) of the system. A knockout of a reaction that participates in 90% of all reaction routes in the system is obviously more severe than that of a reaction being part in 10% of all reaction routes. These bottlenecks can be found out very easily by sorting the ratios $z^{(i)}/z$, with $z^{(i)}$ being the number of remaining EFM after knockout i and z being the total number of EFMs, in an ascending order. The knockouts with the lowest numbers of remaining EFMs are then at the top of the list. I show this approach below in the Section "Structural robustness based on multiple knockouts" for the four metabolic networks analysed in our second publication (Behre *et al.*, 2008, Chapter 3).

To cope with subsets of EFMs producing essential metabolites, we introduced the robustness measures R_2 and R_3. Whereas R_2 is defined as the minimum of all particular product robustnesses, the measure R_3 deals with cases where the chosen metabolites are important for the organism but not entirely vital. If these metabolites are of different importance, also weighted mean of the particular product robustnesses can be used. The weighting factors can, for instance, be derived from the numbers of EFMs producing the concerning product.

The definitions introduced in our first paper (Wilhelm *et al.*, 2004) essentially take into account single mutants. Nevertheless, depending on the level of redundancy, there might be two or more genes (or enzymes) to be knocked out to block a certain (essential) route in metabolism. For instance, in the case of *synthetic lethal mutations*, it needs two knockouts at the same time to block a certain cell function. As also mentioned in Deutscher *et al.* (2008), on the basis of single knockouts, it could not be distinguished between two metabolic functions being completely superfluous and two metabolic functions being completely redundant. Consequently, the smallest MCS (Klamt and Gilles, 2004; Klamt, 2006) of such a system comprises more than one enzyme (or gene). Hence, it is important to analyse also the behaviour of a system in the case of double or multiple knockouts. To cope with this necessity, we extended our robustness framework to double or multiple knockouts.

Furthermore, the differences in structural robustness between different

organisms (and also between different parts of metabolisms in one organism) raise questions concerning the evolutionary behaviour of structural robustness. Thus, in further studies, it will be of interest to analyse the change of robustness of metabolism during biological evolution (cf. Ebenhöh and Heinrich, 2003). Since evolution usually prefers traits enhancing the robustness of an organism (Kitano, 2004, cf. General Introduction, Section 1.3), one would assume an increase in robustness. Nevertheless, also the opposite change can happen as well. This is due to the effect that an increase in robustness is always accompanied by a rise in metabolic effort. Therefore, also a loss in metabolic robustness can occur if the environmental conditions become more convenient, as it is the case for erythrocytes or, for example, during the evolution of intracellular parasites. In this context, it is worth to further study the evolution of enzymes with broad substrate specificity. It has been argued that highly specific enzymes have developed from ancestors with low-specificity during biological evolution (Kacser and Beeby, 1984). An explanation for this development may be an increase in robustness, since several specialised enzymes require more knockouts to block all their functions than one less specific enzyme.

Structural robustness based on multiple knockouts

In our second publication (Behre *et al.*, 2008, Chapter 3), we generalised the robustness measures for metabolic networks, taking into account single, double, and multiple knockouts. These measures $R_1(d)$, $R_2(d)$, and $R_3(d)$ with d being the knockout depth are analogously defined as their equivalents for single knockouts, just that they are based on the ratio of the number of EFMs in the unperturbed situation in comparison to the number of remaining EFMs after knockout of *one or several enzymes*, averaged over all *combinations of knockouts*. As for single knockouts, this results in normalised quantities that do not depend on the size of the networks. Moreover, as already demonstrated for single knockouts, the robustness values are only based on the network topology, independently of the function of the system under study. First, with the help of simple examples (Table 2 in Behre *et al.*, 2008), we evaluated the behaviour of metabolic networks against knockouts

with different cardinalities d. We could generally prove the intuitively comprehensible observation that the robustness decreases if the cardinality d of knockouts increases. The examples furthermore show that for given numbers of enzymes and EFMs the robustness against double or multiple knockouts is higher if two branches in the network have different lengths (cf. systems 1 and 2 in Table 2 in Chapter 3). It is interesting that such cases often occur in metabolism, for instance, the different amino acids are synthesised on pathways of very different lengths. Probably, chemical constraints are the main reasons for this observation. Nevertheless, it might be that robustness issues also play a role in metabolic network evolution. Our small examples also demonstrate that the position of a branching point in metabolism is of importance for the structural robustness. Assumed the case that two pathways of same lenght are compared, a branching point that is more downstream (regarding to the metabolic flux) results in a longer bottleneck upstream of branching point and vice versa.

As already mentioned in Section "Structural robustness based on single knockouts", to analyse the fraction $z^{(i)}/z$ of remaining EFMs after knockout of reaction E_i separately for all blockable reactions in the system under study is a good way to determine bottlenecks of this system. If, for instance, after knockout of reaction "X" only 10% of all EFMs remain, this reaction is obviously more essential than reaction "Y", after the knockout of which 60% of all EFMs are still feasible. Reactions forming bottlenecks can be found very easily by sorting the ratios $z^{(i)}/z$, with $z^{(i)}$ being the number of remaining EFMs after knockout i and z being the total number of EFMs, in an ascending order. The knockouts with the lowest numbers of remaining EFMs are then listed at the top of the table. Here, I show this approach for the metabolic networks analysed in our second publication (Behre *et al.*, 2008, Chapter 3). In Table 5.1, I show the ten most essential reactions for the entire amino acid anabolism in *Escherichia coli* (Eco), the part of the amino acid anabolism in *Escherichia coli* that just comprises those amino acids that are essential for humans (Eco_es), the part of the amino acid anabolism in *Escherichia coli* that just comprises those amino acids that are non-essential for humans (Eco_nes), and the amino acid anabolism in human hepatocytes (Hsa).

Table 5.1. The ten most essential reactions for the entire amino acid anabolism in *Escherichia coli* (Eco), the part of the amino acid anabolism in *Escherichia coli* that just comprises those amino acids that are essential for humans (Eco_es), the part of the amino acid anabolism in *Escherichia coli* that just comprises those amino acids that are non-essential for humans (Eco_nes), and the amino acid anabolism in human hepatocytes (Hsa)

Eco		**Eco_es**		**Eco_nes**		**Hsa**	
Reaction	rem. EFMs	Reaction	rem. EFMs	Reaction	rem. EFMs	Reaction	rem. EFMs
R01324	2,0%	R01082	6,9%	R01324	1,7%	Pyr_up	7,7%
R00355	7,6%	R02164	6,9%	R00268	9,6%	R00209	12,0%
R01082	16,0%	Ery_up	18,5%	R01899	9,6%	R00344	12,0%
R00586	17,4%	R01714	18,5%	R01082	17,9%	R00351	12,0%
R00897	17,4%	R01826	18,5%	SufS	23,8%	R01324	12,0%
R00480	20,4%	R02412	18,5%	R00258	27,4%	OGC	23,2%
R02291	20,4%	R02413	18,5%	R00355	31,1%	R00782	25,4%
R01773	22,3%	R03083	18,5%	G3P_up	31,5%	R00342	27,1%
R00268	22,9%	R03084	18,5%	R00582	31,5%	DIC_1	33,8%
R01899	22,9%	R03460	18,5%	R01061	31,5%	Pro_tr	34,4%

As can be seen from Table 5.1 the most vulnerable points in the entire amino acid anabolism in *Escherichia coli* (Eco) are the reactions R01324 and R00355, catalysed by the enzymes "Citrate(isocitrate) hydro-lyase (cis-aconitate-forming)" and "L-aspartate:2-oxoglutarate aminotransferase", respectively. "Citrate(isocitrate) hydro-lyase (cis-aconitate-forming)" is one of the enzymes of the citrate cycle (TCA cycle) and thus of central importance for the energy supply of anabolic reactions. "L-aspartate:2-oxoglutarate aminotransferase" converts Oxaloacetate to L-Aspartate. Hence, this enzyme is on the one hand closely related to the citrate cycle and on the other hand essential for L-Aspartate synthesis. The next reaction R01082 is catalysed by the enzyme "(S)-malate hydro-lyase (fumarate-forming)" and thus also an essential part of the TCA cycle. From the remaining seven reactions five (R00586, R00897, R00480, R02291, and R01773, for full enzyme names see Glossary) belong to the cysteine and methionine metabolism and two (R00268 and R01899) again to the citrate cycle. For the part of the amino acid anabolism in *Escherichia coli* that just comprises those amino acids that are essential for humans (Eco_es)

the reactions R01082 ("(S)-malate hydro-lyase (fumarate-forming)") and R02164 ("Succinate:ubiquinone oxidoreductase"), both from the TCA cycle are the most important steps in the network. The third reaction (Ery_up) is an artificial uptake reaction for D-Erythrose 4-phosphate that replaces the pentose phosphate pathway. This pathway was not fully included into the amino acid anabolism (together with glycolysis) to focus more on the amino acid synthesis. But the importance of this artificial uptake reaction shows the significance of the pentose phosphate pathway. In the other part of the amino acid anabolism in *Escherichia coli* comprising those amino acids that are non-essential for humans (Eco_nes) the reaction R01324 ("Citrate(isocitrate) hydro-lyase (cis-aconitate-forming)") is the Achilles heel (as in the entire network) followed by R00268, and R01899 both catalysed by the enzyme "Isocitrate:NADP+ oxidoreductase (decarboxylating)". Accordingly, all three reactions are part of the TCA cycle. In the amino acid anabolism in human hepatocytes (Hsa) the artificial uptake reaction for Pyruvate that replaces glycolysis is the most crucial point in the network. This shows very impressively the significance of the glycolysis. The next reaction is R00209, the mitochondrial multienzyme pyruvate dehydrogenase complex consisting of the enzymes "Pyruvate:[dihydrolipoyllysine-residue acetyltransferase]-lipoyllysine 2-oxidoreductase (decarboxylating, acceptor-acetylating)" (EC 1.2.4.1), "Acetyl-CoA:enzyme N6-(dihydrolipoyl)lysine S-acetyltransferase" (EC 2.3.1.12), and "Protein-N6-(dihydrolipoyl)lysine:NAD+ oxidoreductase" (EC 1.8.1.4), that catalyses the conversion from pyruvate to acetyl coenzyme A and thus connects glycolysis with the citrate cycle. At the third position is the reaction R00344 ("Pyruvate:carbon-dioxide ligase (ADP-forming)") that is closely related to citrate cycle. This reaction converts ATP-driven pyruvate to oxaloacetate. The results in Table 5.1 show very impressively how crucial parts of metabolic networks can be determined with the help of EFMs.

In our previous work (Wilhelm *et al.*, 2004) enzymes were lumped to some extent. A comparison of examples 3 and 5 in Table 2 (Chapter 3) shows that a lumping of enzymes affects the robustness values, even if the combined enzymes are part of the same enzyme subset (cf. Pfeiffer *et al.*,

1999). Thus, we have here considered each enzyme separately. In metabolic modelling, isoenzymes are often lumped into combined reactions. Whereas this is adequate for many applications, it is not when robustness to enzyme deletions is studied, because no particular reactions but particular enzymes are knocked out. Thus, for instance, succinate dehydrogenase and fumarate reductase in *Escherichia coli* need to be distinguished. Analogously, enzymes with broad substrate specificity require special attention. The knockout of an enzyme catalysing several reactions implies the deletion of all these reactions (unless they are catalysed by other enzymes simultaneously).

The overall robustness that we introduced as a weighted sum of all robustnesses $R_1(d)$ up to a certain cardinality d, lies for all considered systems between the robustness against single knockouts and that against double knockouts (Tables 3-5 in Behre *et al.*, 2008). Figure 1 in Chapter 3 shows that it is not necessary to determine robustness for cardinalities larger than 3 since the robustness values approach an asymptotic lower limit. On one hand this is plausible, since the probability of a knockout to occur decreases if the knockout depth increases. On the other hand, this is also of advantage, since the question arises whether our robustness approach is scalable to larger, e.g. genome-wide networks. In this sense, our robustness approach is of limited applicability. One reason is that already the number of EFMs grows exponentially with the network size. And furthermore, with increasing knockout depth we deal with a combinatorial explosion of knockouts. However, on one hand, a cardinality of 3 is already sufficient to estimate the overall robustness. And on the other hand, there are already approaches to calculate at least subsets of EFMs in genome-scale networks using linear programming (de Figueiredo *et al.*, 2009b; Kaleta *et al.*, 2009).

There is an interesting similarity between our equation to calculate structural robustness (Wilhelm *et al.*, 2004; Behre *et al.*, 2008) and a more general one that was proposed by Kitano (2007). He suggested the following formula to estimate the robustness R of a system s regarding a function a against a set of perturbations P as:

$$R^s_{a,P} = \int_P \psi(p) D^s_a(p) dp \tag{5.1}$$

with $\psi(p)$ being the probability for perturbation p to take place, $D_a^s(p)$ being an evaluation function for perturbation p, and P being the entire perturbation space. Since Kitano considers all kinds of perturbations (not just knockouts as we do) his perturbation space can be continuous whereas ours is discrete. Kitano's evaluation function determines whether the system still maintains its function in case of perturbation p and to what degree, respectively. In our equation (see below the generalised version Eq. 5.2) the perturbation space is the set of all knockout combinations with cardinality d from r blockable reactions; and since this set contains discrete entries we perform a summation rather than integration over the perturbation space.

$$R_1(d) = \frac{\sum_{i=1}^{\binom{r}{d}} z^{(i)}}{\binom{r}{d} \cdot z} \tag{5.2}$$

In our formula, that can also be written as Eq. 5.3, the ratio $1/\binom{r}{d}$ can be considered as a conditional probability. Given that one knockout with cardinality d occurs, the probability that one of $\binom{r}{d}$ knockout combinations is hit, is $1/\binom{r}{d}$.

$$R_1(d) = \sum_{i=1}^{\binom{r}{d}} \frac{z^{(i)}}{\binom{r}{d} \cdot z} \tag{5.3}$$

The ratio $z^{(i)}/z$ can be seen as a roughly simplified evaluation function. Given, for instance, a certain knockout with cardinality d that blocks 50% of all EFMs, the system is considered to maintain 50% of its original functionality. From that point of view, Kitano's equation from 2007 can be considered as a generalised version of our ealier proposed robustness measure.

Since in the course of evolution metabolic functions get lost that are no longer essential due to more comfortable environmental conditions (Soyer and Pfeiffer, 2010), it is plausible to argue that those functionalities disappear that are stronger affected by knockouts, for example due to a lower interconnectivity in the metabolic network. A good example in this sense is the amino acid anabolism. For humans, the amino acids histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine are essential, whereas, for instance, for *Escherichia coli* they are not. Al-

though tyrosine is considered as non-essential in many textbooks because it can be synthesised from phenylalanine (and only from phenylalanine) in one step, we classify it as essential (in agreement with Voet and Voet, 2004) because phenylalanine is essential. In the course of its evolution, *Homo sapiens* developed from ancestors that were pure herbivores to an omnivorous species. Since the supply with proteins and thus amino acids is much more ensured by herbal and animal nutrition, than by purely herbal nutrition, the evolutionary imperative to synthesise all amino acids was no longer given. Hence, mutations disabling the metabolism of rarely used amino acids were no longer strictly lethal. We therefore applied our extended framework to metabolic networks representing amino acid anabolism in *Escherichia coli* and human hepatocytes, and moreover to the central metabolism in human erythrocytes. We subdivided the *Escherichia coli* model into two subnetworks that were studied separately: (i) the subnetwork synthesising amino acids that are essential for humans, and (ii) the subnetwork synthesising amino acids that are non-essential for humans (`Metatool` models of these networks see Appendix A).

As expected, among the analysed networks the erythrocyte model shows the lowest robustness, which is in agreement with the results of our first publication (Wilhelm *et al.*, 2004). The hepatocyte metabolism is more robust than erythrocyte metabolism, but less robust than the *Escherichia coli* metabolism. This result is intuitively comprehensible, since erythrocytes must be as small as possible in order to pass thin capillaries. In addition, they are densely packed with haemoglobin for oxygen transport. Hence, only the most necessary parts of metabolism have been maintained in evolution. As erythrocytes, also hepatocytes live under homeostatic conditions. Nevertheless, due to the many different functions of the liver, hepatocytes need to have much more diverse metabolic capabilities. In contrast, *Escherichia coli* must adapt to many different environmental conditions. Thus its metabolism needs to be even more robust than that of hepatocytes. As a consequence, *Escherichia coli* synthesises all amino acids while hepatocytes (as all human cells) save the metabolic costs for producing those amino acids being essential for human. Another reason for the higher robustness of *Escherichia coli* is that the compartmentation in hepatocytes implies transporters form-

ing bottlenecks in the metabolic system. Since not every metabolite can cross intracellular membranes, it can be assumed that compartmentation of networks that are not already modularised reduces structural robustness in many cell types.

Our robustness analysis of the two subnetworks of the *Escherichia coli* amino acid anabolism provides very interesting results. The subnetwork producing those amino acids that are essential for humans, is less robust than that producing non-essential amino acids. This encourages the hypothesis that the non-essential amino acids are earlier adopted in evolution and thus more cross-linked in the metabolic network, since evolution prefers traits enhancing the robustness, which on the structural level of a metabolic network can be achieved by a higher interconnectivity. In contrast, the amino acids that are essential for humans, might be accrued later in evolution (Jordan *et al.*, 2005), and thus their anabolic pathways are less interconnected. This lower interconnectivity might be the reason why their synthesis pathways got lost in the evolution towards higher organisms such as humans. Nevertheless, these structural differences are still visible in organisms like *Escherichia coli* that are still able to sythesise all amino acids. To further confirm this hypothesis, further amino acid anabolisms need to be compared.

The structural background for these differences in robustness is also visible by manual inspection. The synthesis pathways of essential amino acids such as tryptophan or isoleucine are relatively "straight", meaning that they do not involve many branching points. Moreover, in the case of prokaryotes, the corresponding genes are often regulated as operons (e.g. the tryptophan operon), so that a mutational loss of a whole pathway can occur very easily. In contrast, the synthesis of the non-essential amino acids is based on pathways with a higher level of ramification and is entangled in the synthesis of other compounds, which results in much more redundancy.

Now the question arises, how Jordan *et al.* (2005) estimated the evolutionary age of amino acids. Their hypothesis is that amino acids developed later in the course of evolution are statistically still underrepresented. Thus, they assume that mutations more often lead to an increased use of those amino acids than to a decreased use. Their results are quite amazing. They compared sets of orthologous proteins encoded by triplets of closely related

genomes from 15 taxa representing all three domains of life (*Bacteria*, *Archaea*, and *Eukaryota*), and used phylogenetic methods to point out amino acid substitutions. For histidine, isoleucine, lysine, methionine, phenylalanine, tryptophan, and valine, Jordan *et al.* (2005) found more mutations leading to a more frequent usage than mutations leading to a less frequent usage. All these amino acids are essential for humans. In contrast, amino acids like alanine, glutamic acid, glycine, or proline that are all non-essential for humans, are constantly lost. Jordan *et al.* (2005) assume that all amino acids with declining frequencies are incorporated at first into the genetic code, whereas those with increasing frequencies were probably recruited later. Thus, the expansion of initially underrepresented amino acids, which began over 3,400 million years ago, obviously proceeds to this day. These results are an indirect evidence for the hypothesis that in the course of evolution the robustness of a metabolic network increases (if necessary) by enhancing the interconnectivity of the network. This implies that metabolites adopted later in evolution are less interconnected. If now, as in the case of the essential amino acids, the evolutionary pressure to assure a certain part of metabolism declines, the corresponding metabolites may also faster disappear again.

For applications in pharmacology and biotechnology, it is always important to find weaknesses in metabolisms or signalling systems of pathogens, since the mechanism of antibiotics consists of being lethal for pathogens but not for human cells by exploiting Achilles heels of pathogen metabolism that are not existing in human cells (Lüllmann *et al.*, 2006). In this sense, our robustness approach is important for future developments in pharmaceutical industries, since it provides a compact way to assess the robustness of different metabolisms. Since the progress in detecting drugs acting on single proteins has slowed down (Huang, 2001; Frantz, 2005), drugs combining two or more enzyme inhibitors have recently attracted increasing interest. Such adjunctions are of interest in treating bacterial infections (Barchiesi *et al.*, 2004), AIDS (Taburet *et al.*, 2004), and others. A similar situation exists in biotechnology. Here, the suppression of inefficient pathways often requires the additional deletion of undesired side reactions, so that multiple knockouts become necessary (Carlson *et al.*, 2002; Vijayasankaran *et al.*, 2005; Trinh *et al.*, 2008, 2009; Trinh and Srienc, 2009).

HepatoNet1: reconstruction of liver metabolism

The `Metatool` model of the amino acid anabolism of hepatocytes that I compiled for our robustness studies in Behre *et al.* (2008) is listed in the Appendix of this thesis (see Appendix A.4). It comprises 82 reactions including six spontaneous reactions and—since we considered compartmentation by distinguishing between reactions in the cytosol and those in mitochondria—also 13 exchange reactions between these two compartments. This model gives rise to 712 EFMs. During literature search I had to do for the compilation of our model, I found 25 further exchange reactions comprising both diffusion and active transporters like symports and antiports. These additional transport reactions are listed in Appendix A.5. I contributed our hepatocyte amino acid anabolism and the additional transport reactions to the comprehensive hepatocyte metabolism *HepatoNet1* that was reconstructed by the group of Prof. Holzhütter at the Charité in Berlin (see Gille *et al.*, 2010, where I am also a co-author).

Together with Jerby *et al.* (2010), HepatoNet1 is the first reconstruction of a comprehensive metabolic network of the human hepatocyte accomplishing a large number of known metabolic liver functions. The network comprises 777 metabolites within six intracellular and two extracellular compartments. It contains 2539 reactions, including 1466 transport reactions and is based on the manual appraisal of more than 1500 original scientific publications. The final network was compiled with the help of an iterative process of data assembly and computational testing of network functionalities with constraint-based modelling techniques. During the curation process the integration of data from different sources was managed with the program `METANNOGEN` that was programmed for this purpose (see Gille *et al.*, 2007). The detoxification of ammonia in hepatocytes was used as an example to show the efficient response of the liver to perturbations of the homeostasis of blood compounds.

Since the objective was to compile a functionally comprehensive network rather than a collection of all reactions and metabolites referred for hepatocytes, the model was intensively tested using FBA. For these analyses 442 different metabolic objectives were prepared that had to be accomplished,

that is the network model had to determine a non-zero stationary flux distribution. Additionally, it was also checked that biochemically impossible tasks are not realisable with the model. The computations were carried out using the CPLEX package (http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/). Resulting flux modes were visualised either with the program BiNA 1.3.1 (Küntzer *et al.*, 2007) or with CytoScape (Shannon *et al.*, 2003) in combination with FluxViz (König and Holzhütter, 2010). The flux-balance computations were executed by identifying stationary flux distributions that as well minimise the sum of internal network fluxes (Holzhütter, 2004, 2006) as obey the criterion of thermodynamic realisability, meaning that the directions of the fluxes coincide with the Gibb's free energy of the reactions (Hoppe *et al.*, 2007). For this aim, standard Gibbs energies have been predicted by a method proposed by Jankowski *et al.* (2008) and physiological metabolite concentrations were taken from the *Human Metabolome database* (Wishart *et al.*, 2007). Unreachable reactions that could not be forced to carry any non-zero flux in any tested metabolic objective, were pruned to ensure that the model does not contain "dead parts" (Hoffmann *et al.*, 2007).

An important aspect to explore is the robustness of metabolic liver functions against enzyme deficiencies. Since HepatoNet1 was validated with the help of FBA it is obvious also to estimate the essentiality of enzymes and transporters by using this technique. In the group of Prof. Holzhütter they analysed how knockouts of single enzymes or transporters impair the metabolic objectives accomplished by the human hepatocyte. They performed 123 computational knockout studies, one for each functional flux mode related to the metabolic objectives defined for testing HepatoNet1. In each knockout study the reactions participating in the respective flux mode were blocked one after the other. If such a knockout could be compensated by an alternative flux mode the corresponding was defined to be non-essential. Since the essentiality of enzymes and transporters depends also on the presence of external substrates that can be used to bypass the impaired reaction, the knockout studies were conducted with less restrictive exchange sets. Accordingly, these knockout simulations output less essential enzymes than at minimal conditions, but certainly higher numbers than at comfortable con-

ditions when hepatocytes have full access to all metabolites in the plasma.

Similar to our proposal to sort the fractions $z^{(i)}/z$ of remaining EFMs, discussed in Section "Structural robustness based on single knockouts" and shown in the Section "Structural robustness based on multiple knockouts" for the four metabolic networks analysed in our second publication (Behre *et al.*, 2008, Chapter 3), also here the essential reactions ascertained in the 123 knockout studies were ranked. But in contrast to our ranking here a descending order according to the frequency of their occurrence was chosen. Essential reactions at the top of this ranking were defined to be cardinally essential, that is they are indispensable for almost all tested metabolic objectives (e.g. electron carriers of the respiratory chain). Non-essential reactions, in turn, were listed at the last positions. An intriguing result of these knockout studies is that there are not only reported enzymopathies with clinical symptoms for those enzymes and transporters that turned out to be weakly essential, but also for strongly essential ones, which in case of a complete knockout are predicted to impair a larger set of metabolic objectives, and thus are probably lethal. An explanation could be the existence of protein isoforms (not included in the model) that are not affected by the specific type of deficiency.

The method of choice for validating the HepatoNet1 model is FBA. A reason for that choice is on one hand the possibility to introduce more constraints than the set of nutrients in the environment (here the plasma). The other reason is that the HepatoNet1 model is too large to calculate all EFMs. Thus, it is not possible to determine a robustness value with our method, even not for single knockouts. It would be interesting to compare the robustness that we calculated for our hepatocyte amino acid anabolism with the corresponding one for HepatoNet1. But since HepatoNet1 is a whole-cell model whereas our hepatocyte amino acid anabolism is just a cut-out of it and thus—as already discussed in the Section "Structural robustness based on single knockouts"—less interconnected, it can be suggested that the robustness of HepatoNet1 is clearly higher than that of our amino acid anabolism.

Detecting signalling pathways in enzyme cascades

In our third publication (Behre and Schuster, 2009, Chapter 4), we have presented an approach for detecting elementary signalling routes in enzyme cascades, such as phosphorylation cascades. We adapted EFM analysis established earlier for detecting pathways in metabolic networks (Schuster *et al.*, 1999, 2000a).

A strict application of the concept of EFMs to enzyme cascades would lead to trivial results. Rather than the routes of information transfer particular enzyme cycles (e.g. phosphorylation-dephosphorylation cycles) would be reflected by EFMs. To describe information-transfer routes in an applicable way, we argue that signalling usually involves signal amplification or at least constant signal strength, but never diminuition. Thus, in the limit case of constant signal strength a minimum stoichiometry is given: each active enzyme molecule needs to activate at least one enzyme molecule at the next level of the cascade in order not to disrupt the flux of information. Furthermore, we assume a pseudo-steady state, since a signalling system has to return to its initial state before the next signalling event occurs. Thus, averaging over the time span needed for the signalling event and the recoverage period, the concentrations of the involved components stays constant. Moreover, from a chronological point of view, the phosphorylation and dephosphorylation processes are interlocked along a cascade. Let us start with the phosphorylation of kinase E_2 by kinase E_1P. Next, the dephosphorylation of E_1P rises while E_2P is already phosphorylating E_3 and so on. Hence, we write reaction equations where phosphorylation of kinase E_{n+1} is coupled with dephosphorylation of kinase E_n. These coupled reaction equations lead to "EFMs" representing routes of information transfer along the cascade. I quote the term "elementary flux modes" here, since the flow of information just coincides with a seeming mass flow. Hence, it is better to call such routes "elementary signalling pathways" or "elementary signalling modes" (ESMs). From the overall reaction equation of such a pathway one can see that not one phosphate is passed along the cascade, but rather the phosphate is renewed at each cascade level. Therefore, we consider such enzyme cascades as relay races with changing batons. This situation

differs from group-transfer pathways such as the phosphotransferase system (PTS) for which the analogy to a relay race is more appropriate because the “baton” is not replaced. Papin and Palsson (2004) studied this type of systems. In the case of signal amplification the same routes are relevant. The level of amplification can be expressed by an appropriate stoichiometry. Assumed that on average 10 molecules E_2 are phosphorylated by E_1P before E_1P gets dephosphorylated, the corresponding reaction equation would be $E_1P + 10\,E_2 + 10\,NTP = E_1 + P + 10\,E_2P + 10\,NDP$.

By this reasoning we have put the application of EFM analysis and related methods like extreme pathway analysis (Schilling *et al.*, 2000) and minimal T-invariants (cf. Starke, 1990) to intracellular signalling systems on a firm theoretical basis. Earlier, these analyses have been applied in a formal way to such systems without a theoretical justification (Xiong *et al.*, 2004; Heiner *et al.*, 2004; Sackmann *et al.*, 2006).

To demonstrate our approach we used prototypic single and double phosphorylation cascades. Nevertheless, our approach works with all chemical modifications that switch the state of an enzyme from active to inactive and back. In our paper (Behre and Schuster, 2009), we mentioned also methylation, acetylation, adenylylation, and ubiquitination. But, as can be seen from the part of the insulin signalling network that we applied our approach to, the switch from active to inactive can also be induced by reversible complex formations. Another example is the enzyme Ras, which is included in our model. It belongs to the family of small GTPases. Such enzymes are activated by replacing the GDP they are complexed with in their inactive form, by GTP. This substitution is in fact equivalent to a phosphorylation. Since also the phosphorylations can occur in different ways depending on the types of phosphate donors and acceptors (ATP/ADP, GTP/GDP etc.), we used the unspecific symbols NTP and NDP.

Our insulin signalling network was compiled on the basis of data taken from TRANSPATH® database (Krull *et al.*, 2006). Its complex formations were modelled as usual metabolic reactions. But at the switch to the next level of the cascade the complexes are decomposed into their original components to fulfil the assumed pseudo-steady state. The three resulting ESMs represent three different biochemical functions of the insulin sig-

nalling network. The first ESM leads to a serine phosphorylated and thus active 6-phosphofructo-2-kinase / fructose 2,6-bisphosphatase (EC 3.1.3.46 / EC 2.7.1.105), a bifunctional enzyme participating in fructose and mannose metabolism and catalysing the reaction from β-D-fructose 6-phosphate to β-D-fructose 2,6-bisphosphate and back. The second ESM results in an activated transcription factor CREB(PS) that regulates the expression of the insulin-like growth factor binding protein 1, which leads to regulation of cell growth, glucose metabolism and several other processes. The third ESM provides a phosphorylated ribosomal protein S6 finally inducing activation of protein synthesis (for details see the TRANSPATH® database).

The entire insulin signalling network comprises many more functions than the part that we have chosen, such as activation of glycogenesis and enhancing glucose uptake in muscle cells and adipocytes by increasing the number of GLUT4 transporters in their membranes. Beside the intention to model a clear example of a real kinase cascade there is one more reason not to model the entire insulin signalling network: With our approach it is not possible to model two (or more) consecutive inhibitions in the same way as it can be achieved within a Boolean framework. In a Boolean model, for instance two consecutive inhibitions can be "summed up" to an activation. But this logical result does no correspond with the "molecular reality". On the molecular level the inhibitor I_1 for enzyme E_1 is, for instance, inhibited by forming a complex with another inhibitor I_2. Hence, no I_1 is left anymore to suppress the activity of E_1. But independent of whether E_1 is inhibited or not, E_1 must first of all exist. And its existence is a result of the preceding transcription and translation. Thus, for staying close to the molecular level, two consecutive inhibitions should possibly be modelled as an ESM ending with an inhibited I_1 (set as external), followed by a new ESM starting with an external precursor of E_1 that represents transcription and translation.

Outlook

The results presented here raise new questions. First of all, the hypothesis that the evolutionary age of different parts of the amino-acid metabolism correspond with the level of interconnectivity between the accordant bio-

chemical reactions needs to be further examined. Moreover, the possibilities to further extend the concept of EFMs to signalling and gene-regulatory networks need to be explored in more detail. The concept of EFMs is a well established framework with a lot of applications based on it, besides our concept of structural robustness (Wilhelm *et al.*, 2004; Behre *et al.*, 2008), for instance, the concept of MCSs (Klamt and Gilles, 2004; Klamt, 2006) or the concept of enzyme subsets (Pfeiffer *et al.*, 1999) and the related concepts in flux coupling analysis (Burgard *et al.*, 2004). A full extension of the concept of EFMs to signalling and gene-regulatory networks would allow for an application of all abovementioned tools to these network types in a straightforward manner.

A converse approach is to elucidate the feasibilities to adopt these tools to other concepts of structural network analysis. For instance, our measures of structural robustness can be easily applied to T-invariants within the theory of Petri nets or to extreme pathways, since these types of pathways are analogous and very similar to EFMs, respectively. Also signalling pathways calculated on the basis of Boolean networks, where information transfer is analysed without explicit consideration of mass balance constraints, can be used to determine the structural robustness of signalling networks from an information-processing perspective. However, a precondition for this application is that the calculated pathways represent the network topology in a comparable manner as EFMs do. Consequently, the signalling pathways that are calculated with `CellNetAnalyzer` are not suitable, since they are determined on the basis of interaction graphs and thus do not take into account AND-connected interactions (cf. Subsection 1.2.6). It still needs to be evaluated whether pathways detected by the Matrix-Formalism approach of Gianchandani *et al.* (2006) are comparable to EFMs in this sense.

Nevertheless, the application of our concept of structural robustness to signalling pathways will also require several modifications to the concept itself. First of all, in pathways consisting of interactions, no longer the edges but the nodes should be knocked out. In metabolic networks, the edges represent the reactions that are catalysed by enzymes that can be blocked by inhibitors or suppression of transcription and/or translation. If reactions are spontaneous, they cannot be blocked. The nodes of metabolic networks rep-

resent metabolites that could only be blocked by removal, which is (at least *in vivo*) difficult to achieve. In contrast, in signalling networks the nodes are proteins, enzymes, receptors, transcription factors etc., whereas the edges are interactions like complex formations that will (spontaneously) occur if the corresponding components (nodes) are present. Certainly, in the case of phosphorylations and other covalent modifications, catalysing enzymes are involved. But they are also represented as nodes in the network. Hence, in signalling pathways, no longer edges should be considered as preferred targets but rather nodes. Furthermore, if nodes are metabolites like glucose or glycogen, or second messengers like cAMP or PIP_3, they can probably not be blocked and should therefore not be considered as targets for perturbations.

A second modification concerns the kind of perturbations. Metabolic networks are usually impaired by absent or blocked enzymes (knockouts). In contrast, in signal-transduction and gene-regulatory networks also inhibitions can occur, and consequently no longer knockouts are the only way to block certain functions of a system. Also additional inhibitions ("knock-ins") can achieve this intention (cf. the extension of MCSs to MISs in Klamt *et al.*, 2006). Moreover, whereas in metabolic networks an EFM drops out independently of being hit by one or more knockouts, in signalling routes, consisting of activations and inhibitions, two or more perturbations might cancel each other out. Hence, while determining the structural robustness of signalling networks with respect to multiple perturbations, those combinations of disturbances having in total no effect, need to be excluded.

Bibliography

A. A. Agrawal. Phenotypic plasticity in the interactions and evolution of species. *Science*, 294(5541):321–326, 2001.

R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

U. Alon. *An introduction to systems biology: design principles of biological circuits.* Chapman and Hall/CRC Mathematical & Computational Biology, 2007.

S. S. Andrews and D. Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Physical Biology*, 1(3-4):137–151, 2004.

A. Atlante, S. Passarella, P. Pierro, C. Di Martino, and E. Quagliariello. The mechanism of proline/glutamate antiport in rat kidney mitochondria. Energy dependence and glutamate-carrier involvement. *European Journal of Biochemistry*, 241(1):171–177, 1996.

N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625, 2004.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

A.-L. Barabási and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

F. Barchiesi, E. Spreghini, A. M. Schimizzi, M. Maracci, D. Giannini, F. Carle, and G. Scalise. Posaconazole and amphotericin B combination

therapy against *Cryptococcus neoformans* infection. *Antimicrobial Agents and Chemotherapy*, 48(9):3312–3316, 2004.

D. A. Beard, S. dan Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophysical Journal*, 83(1):79–86, 2002.

J. Behre and S. Schuster. Modeling signal transduction in enzyme cascades with the concept of elementary flux modes. *Journal of Computational Biology*, 16(6):829–844, 2009.

J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, and S. Schuster. Structural robustness of metabolic networks with respect to multiple knockouts. *Journal of Theoretical Biology*, 252(3):433–441, 2008.

S. L. Bell and B. Ø. Palsson. expa: a program for calculating extreme pathways in biochemical reaction networks. *Bioinformatics*, 21(8):1739–1740, 2005.

J. M. Berg, J. L. Tymczko, and L. Stryer. *Biochemistry.* W. H. Freeman and Company, New York, 6th edition, 2007.

F. T. Bergmann and H. M. Sauro. SBW - a modular framework for systems biology. In *WSC '06: Proceedings of the 38th conference on Winter simulation*, pages 1637–1645. Winter Simulation Conference, 2006. ISBN 1-4244-0501-7.

J. D. Bloom, L. I. Gong, and D. Baltimore. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science*, 328(5983): 1272–1275, 2010.

A. P. Burgard, E. V. Nikolaev, C. H. Schilling, and C. D. Maranas. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Research*, 14(2):301–312, 2004.

J. M. Carlson and J. Doyle. Complexity and robustness. *Proceedings of the National Academy of Sciences of the United States of America*, 99(Suppl 1):2538–2545, 2002.

R. Carlson, D. Fell, and F. Srienc. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnology and Bioengineering*, 79(2):121–134, 2002.

T. Çakir, B. Kirdar, and K. Ö. Ülgen. Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks. *Biotechnology and Bioengineering*, 86(3):251–260, 2004a.

T. Çakir, C. S. Tacer, and K. Ö. Ülgen. Metabolic pathway analysis of enzyme-deficient human red blood cells. *Biosystems*, 78(1-3):49–67, 2004b.

J. S. Clegg. Cryptobiosis–a peculiar state of biological organization. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 128(4):613–624, 2001.

A. Cornish-Bowden. Putting the systems back into systems biology. *Perspectives in Biology and Medicine*, 49(4):475–489, 2006.

A. Cornish-Bowden and M. L. Cárdenas. Systems biology may work when we learn to understand the parts in terms of the whole. *Biochemical Society Transactions*, 33(Pt 3):516–519, 2005.

A. Cornish-Bowden, M. L. Cárdenas, J.-C. Letelier, J. Soto-Andrade, and F. G. Abarzúa. Understanding the parts in terms of the whole. *Biology of the Cell*, 96(9):713–717, 2004.

M. Csete and J. Doyle. Bow ties, metabolism and disease. *Trends in Biotechnology*, 22(9):446–450, 2004.

L. F. de Figueiredo, A. Podhorski, A. Rubio, J. E. Beasley, S. Schuster, and F. J. Planes. Calculating the K-shortest elementary flux modes in metabolic networks. In I. Troch and F. Breitenecker, editors, *Proceedings of the MATHMOD Conference 2009 in Vienna.*, pages 736–747, 2009a.

L. F. de Figueiredo, S. Schuster, C. Kaleta, and D. A. Fell. Can sugars be produced from fatty acids? A test case for pathway analysis tools. (Erratum). *Bioinformatics*, 25(1):152–158, 2009b.

J. L. DeRisi, V. R. Iyer, and P. O. Brown. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338): 680–686, 1997.

D. Deutscher, I. Meilijson, M. Kupiec, and E. Ruppin. Multiple knockout analysis of genetic robustness in the yeast metabolic network. *Nature Genetics*, 38(9):993–998, 2006.

D. Deutscher, I. Meilijson, S. Schuster, and E. Ruppin. Can single knockouts accurately single out gene functions? *BMC Systems Biology*, 2:50, 2008.

B. Di Ventura, C. Lemerle, K. Michalodimitrakis, and L. Serrano. From *in vivo* to *in silico* biology and back. *Nature*, 443(7111):527–533, 2006.

R. Diestel. *Graphentheorie.* Springer-Verlag Berlin and Heidelberg GmbH & Co. KG, Germany, 2000.

P. Dittrich and P. Speroni di Fenizio. Chemical organisation theory. *Bulletin of Mathematical Biology*, 69(4):1199–1231, 2007.

O. Ebenhöh and R. Heinrich. Stoichiometric design of metabolic networks: multifunctionality, clusters, optimization, weak and strong robustness. *Bulletin of Mathematical Biology*, 65(2):323–357, 2003.

J. S. Edwards and B. Ø. Palsson. How will bioinformatics influence metabolic engineering? *Biotechnology and Bioengineering*, 58(2-3):162–169, 1998.

J. S. Edwards and B. Ø. Palsson. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnology Progress*, 16(6):927–939, 2000.

A. M. Feist and B. Ø. Palsson. The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*, 26(6):659–667, 2008.

D. A. Fell and J. R. Small. Fat synthesis in adipose tissue: an examination of stoichiometric constraints. *Biochemical Journal*, 238(3):781–786, 1986.

E. Fischer and U. Sauer. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *Journal of Biological Chemistry*, 278(47):46446–46451, 2003.

S. Frantz. Drug discovery: playing dirty. *Nature*, 437(7061):942–943, 2005.

A. Gevorgyan, M. G. Poolman, and D. A. Fell. Detection of stoichiometric inconsistencies in biomolecular models. *Bioinformatics*, 24(19):2245–2251, 2008.

E. P. Gianchandani, J. A. Papin, N. D. Price, A. R. Joyce, and B. Ø. Palsson. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Computational Biology*, 2(8 e101):0902–0917, 2006.

C. Gille, S. Hoffmann, and H.-G. Holzhütter. METANNOGEN: compiling features of biochemical reactions needed for the reconstruction of metabolic networks. *BMC Systems Biology*, 1:5, 2007.

C. Gille, C. Bölling, A. Hoppe, S. Bulik, S. Hoffmann, K. Hübner, A. Karlstädt, R. Ganeshan, M. König, K. Rother, M. Weidlich, J. Behre, and H.-G. Holzhütter. HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Molecular Systems Biology*, 6:411, 2010.

L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402(6761 Suppl):C47–C52, 1999.

M. Heiner, I. Koch, and J. Will. Model validation of biological pathways using Petri nets–demonstrated for apoptosis. *Biosystems*, 75(1-3):15–28, 2004.

M. Heiner, R. Richter, M. Schwarick, and C. Rohr. Snoopy - a tool to design and execute graph-based formalisms. *Petri Net Newsletter*, 74:8–22, 2008.

R. Heinrich, S. M. Rapoport, and T. A. Rapoport. Metabolic regulation and mathematical models. *Progress in Biophysics and Molecular Biology*, 32 (1):1–82, 1977.

R. Hofestädt. A Petri net application to model metabolic processes. *Systems Analysis Modelling Simulation*, 16(2):113–122, 1994. ISSN 0232-9298.

S. Hoffmann, A. Hoppe, and H.-G. Holzhütter. Pruning genome-scale metabolic models to consistent ad functionem networks. *Genome Informatics*, 18:309–320, 2007.

H.-G. Holzhütter. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *EJB - The FEBS Journal*, 271(14):2905–2922, 2004.

H.-G. Holzhütter. The generalized flux-minimization method and its application to metabolic networks affected by enzyme deficiencies. *Biosystems*, 83(2-3):98–107, 2006.

S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

A. Hoppe, S. Hoffmann, and H.-G. Holzhütter. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Systems Biology*, 1:23, 2007.

S. Huang. Genomics, complexity and drug discovery: insights from Boolean network models of cellular regulation. *Pharmacogenomics*, 2(3):203–222, 2001.

M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novère, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

J. L. Ingraham and F. C. Neidhardt. *Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology.*, volume 1. ASM Press, Washington, DC, 2 volume set edition, 1987.

E. W. Jacobsen and G. Cedersund. Structural robustness of biochemical network models–with application to the oscillatory metabolism of activated neutrophils. *IET Systems Biology*, 2(1):39–47, 2008.

M. D. Jankowski, C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophysical Journal*, 95(3):1487–1499, 2008.

H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

L. Jerby, T. Shlomi, and E. Ruppin. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Molecular Systems Biology*, 6:401, 2010.

I. K. Jordan, F. A. Kondrashov, I. A. Adzhubei, Y. I. Wolf, E. V. Koonin, A. S. Kondrashov, and S. Sunyaev. A universal trend of amino acid gain and loss in protein evolution. *Nature*, 433(7026):633–638, 2005.

A. Joshi and B. Ø. Palsson. Metabolic dynamics in the human red cell. Part I–A comprehensive kinetic model. *Journal of Theoretical Biology*, 141(4): 515–528, 1989.

H. Kacser and R. Beeby. Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. *Journal of Molecular Evolution*, 20(1):38–51, 1984.

H. Kacser and J. A. Burns. The control of flux. *Symposia of the Society for Experimental Biology*, 27:65–104, 1973.

C. Kaleta, F. Centler, and P. Dittrich. Analyzing molecular reaction networks: from pathways to chemical organizations. *Molecular Biotechnology*, 34(2):117–123, 2006.

C. Kaleta, L. F. de Figueiredo, and S. Schuster. Can the whole be less than the sum of its parts? Pathway Analysis in Genome-Scale Metabolic Networks using Elementary Flux Patterns. *Genome Research*, 19(10):1872–1883, 2009.

M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, and Y. Yamanishi. KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, 36(Database issue):D480–D484, 2008.

K. J. Kauffman, P. Prakash, and J. S. Edwards. Advances in flux balance analysis. *Current Opinion in Biotechnology*, 14(5):491–496, 2003.

S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3):437–467, 1969.

S. M. Keating, B. J. Bornstein, A. Finney, and M. Hucka. SBMLToolbox: an SBML toolbox for MATLAB users. *Bioinformatics*, 22(10):1275–1277, 2006.

J. Kielbassa, R. Bortfeldt, S. Schuster, and I. Koch. Modeling of the U1 snRNP assembly pathway in alternative splicing in human cells using Petri nets. *Computational Biology and Chemistry*, 33(1):46–61, 2009.

M. Kirschner and J. Gerhart. Evolvability. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8420–8427, 1998.

H. Kitano. Systems Biology: a brief overview. *Science*, 295(5560):1662–1664, 2002a.

H. Kitano. Computational systems biology. *Nature*, 420(6912):206–210, 2002b.

H. Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, 2004.

H. Kitano. Towards a theory of biological robustness. *Molecular Systems Biology*, 3:137, 2007.

H. Kitano. Violations of robustness trade-offs. *Molecular Systems Biology*, 6:384, 2010.

S. Klamt. Generalized concept of minimal cut sets in biochemical networks. *Biosystems*, 83(2-3):233–247, 2006.

S. Klamt and E. D. Gilles. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, 2004.

S. Klamt and J. Stelling. Combinatorial complexity of pathway analysis in metabolic networks. *Molecular Biology Reports*, 29(1-2):233–236, 2002.

S. Klamt, J. Stelling, M. Ginkel, and E. D. Gilles. FluxAnalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps. *Bioinformatics*, 19(2):261–269, 2003.

S. Klamt, J. Saez-Rodriguez, J. A. Lindquist, L. Simeoni, and E. D. Gilles. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7:56, 2006.

S. Klamt, J. Saez-Rodriguez, and E. D. Gilles. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Systems Biology*, 1:2, 2007.

E. Klipp, W. Liebermeister, C. Wierling, A. Kowald, H. Lehrach, and R. Herwig. *Systems biology, a textbook.* WILEY-VCH-Verlag GmbH & Co. KGaA, Weinheim, Germany, 2009.

I. Koch, B. H. Junker, and M. Heiner. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, 21(7):1219–1226, 2005a.

I. Koch, M. Schüler, and M. Heiner. STEPP–Search Tool for Exploration of Petri net Paths: a new tool for Petri net-based path analysis in biochemical networks. *In Silico Biology*, 5(2):129–137, 2005b.

M. König and H.-G. Holzhütter. FluxViz–cytoscape plug-in for vizualisation of flux distributions in networks. *Genome Informatics*, 24:96–103, 2010.

K. G. Kristinsson. Effect of antimicrobial use and other risk factors on antimicrobial resistance in pneumococci. *Microbial Drug Resistance*, 3 (2):117–123, 1997.

M. Krull, S. Pistor, N. Voss, A. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. Potapov, C. Choi, O. Kel-Margoulis, and E. Wingender. TRANSPATH®: an information resource for storing and visualizing signaling pathways and their pathological aberrations. *Nucleic Acids Research*, 34(Database issue):D546–D551, 2006.

K. K. Kumarasamy, M. A. Toleman, T. R. Walsh, J. Bagaria, F. Butt, R. Balakrishnan, U. Chaudhary, M. Doumith, C. G. Giske, S. Irfan, P. Krishnan, A. V. Kumar, S. Maharjan, S. Mushtaq, T. Noorie, D. L. Paterson, A. Pearson, C. Perry, R. Pike, B. Rao, U. Ray, J. B. Sarma, M. Sharma, E. Sheridan, M. A. Thirunarayan, J. Turton, S. Upadhyay, M. Warner, W. Welfare, D. M. Livermore, and N. Woodford. Emergence of a new antibiotic resistance mechanism in India, Pakistan, and the UK: a molecular, biological, and epidemiological study. *The Lancet Infectious Diseases*, 10 (9):597–602, 2010.

J. Küntzer, C. Backes, T. Blum, A. Gerasch, M. Kaufmann, O. Kohlbacher, and H.-P. Lenhof. BNDB - the Biochemical Network Database. *BMC Bioinformatics*, 8:367, 2007.

P. L. Lakin-Thomas. Transcriptional feedback oscillators: maybe, maybe not... *Journal of Biological Rhythms*, 21(2):83–92, 2006.

K. Lautenbach. Exakte Bedingungen der Lebendigkeit für eine Klasse von Petri-Netzen (in German). *GMD Report*, 82:1–123, 1973.

H. H. Lee, M. N. Molla, C. R. Cantor, and J. J. Collins. Bacterial charity work leads to population-wide resistance. *Nature*, 467(7311):82–85, 2010.

J. M. Lee, E. P. Gianchandani, and J. A. Papin. Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*, 7(2):140–150, 2006.

C. Lemerle, B. Di Ventura, and L. Serrano. Space as the final frontier in stochastic simulations of biological systems. *FEBS Letters*, 579(8):1789–1794, 2005.

B. Lewin. *Genes.* Oxford University Press, 7th edition, 2000.

J. C. Liao, S. Y. Hou, and Y. P. Chao. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnology and Bioengineering*, 52(1):129–140, 1996.

H. Lüllmann, K. Mohra, and M. Wehling. *Pharmakologie und Toxikologie: Arzneimittelwirkungen verstehen - Medikamente gezielt einsetzen.* Thieme-Verlag, Stuttgart, Germany, 2006.

R. Mahadevan and C. H. Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, 5(4):264–276, 2003.

H. Matsuno, A. Doi, M. Nagasaki, and S. Miyano. Hybrid Petri net representation of gene regulatory network. *Pacific Symposium on Biocomputing*, 5:341–352, 2000.

P. Mendes. GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Computer Applications in the Biosciences*, 9(5):563–571, 1993.

M. Morohashi, A. E. Winn, M. T. Borisuk, H. Bolouri, J. Doyle, and H. Kitano. Robustness as a measure of plausibility in models of biochemical networks. *Journal of Theoretical Biology*, 216(1):19–30, 2002.

W. W. Navarre, S. B. Zou, H. Roy, J. L. Xie, A. Savchenko, A. Singer, E. Edvokimova, L. R. Prost, R. Kumar, M. Ibba, and F. C. Fang. PoxA, yjeK, and elongation factor P coordinately modulate virulence and drug resistance in *Salmonella enterica. Molecular Cell*, 39(2):209–221, 2010.

D. Noble. Modeling the heart–from genes to cells to the whole organ. *Science*, 295(5560):1678–1682, 2002.

D. Noble. Systems biology and the heart. *Biosystems*, 83(2-3):75–80, 2006.

B. G. Olivier, J. M. Rohwer, and J.-H. S. Hofmeyr. Modelling cellular systems with PySCeS. *Bioinformatics*, 21(4):560–561, 2005.

M. Pachkov, T. Dandekar, J. Korbel, P. Bork, and S. Schuster. Use of pathway analysis and genome context methods for functional genomics of *Mycoplasma pneumoniae* nucleotide metabolism. *Gene*, 396(2):215–225, 2007.

F. Palmieri. The mitochondrial transporter family (SLC25): physiological and pathological implications. *Pflügers Archiv - European Journal of Physiology*, 447(5):689–709, 2004.

F. Palmieri. Diseases caused by defects of mitochondrial carriers: a review. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1777(7-8):564–578, 2008.

J. A. Papin and B. Ø. Palsson. Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *Journal of Theoretical Biology*, 227(2):283–297, 2004.

J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. Ø. Palsson. Comparison of network-based pathway analysis methods. *Trends in Biotechnology*, 22(8):400–405, 2004.

C. A. Petri. Communication with automata (in German). *Institut für instrumentelle Mathematik, Bonn: Schriften des IIM*, 3:1–132, 1962.

T. Pfeiffer, I. Sánchez-Valdenebro, J. C. Nuño, F. Montero, and S. Schuster. METATOOL: for studying metabolic networks. *Bioinformatics*, 15(3): 251–257, 1999.

M. G. Poolman. ScrumPy: metabolic modelling with Python. *IEE Proceedings - Systems Biology*, 153(5):375–378, 2006.

M. G. Poolman, C. Sebu, M. K. Pidcock, and D. A. Fell. Modular decomposition of metabolic systems via null-space analysis. *Journal of Theoretical Biology*, 249(4):691–705, 2007.

R. K. Porter. Mammalian mitochondrial inner membrane cationic and neutral amino acid carriers. *Biochimica et Biophysica Acta*, 1459(2-3):356–362, 2000.

N. D. Price, J. L. Reed, and B. Ø. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11):886–897, 2004.

T. A. Rapoport, R. Heinrich, G. Jacobasch, and S. Rapoport. A linear steady-state treatment of enzymatic chains. A mathematical model of glycolysis of human erythrocytes. *European Journal of Biochemistry*, 42(1):107–120, 1974.

T. A. Rapoport, R. Heinrich, and S. M. Rapoport. 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. *Biochemical Journal*, 154(15):449–469, 1976.

V. N. Reddy, M. L. Mavrovouniotis, and M. N. Liebman. Petri net representations in metabolic pathways. *Proceedings of the International Conference on Intelligent Systems for Molecular Biology*, 1:328–336, 1993.

C. Rohr, W. Marwan, and M. Heiner. Snoopy–a unifying Petri net framework to investigate biomolecular networks. *Bioinformatics*, 26(7):974–975, 2010.

A. Sackmann, M. Heiner, and I. Koch. Application of Petri net based analysis techniques to signal transduction pathways. *BMC Bioinformatics*, 7:482, 2006.

C. Salazar and T. Höfer. Multisite protein phosphorylation–from molecular mechanisms to kinetic models. *FEBS Journal*, 276(12):3177–3198, 2009.

J. M. Savinell and B. Ø. Palsson. Network analysis of intermediary metabolism using linear optimization: I. development of mathematical formalism. *Journal of Theoretical Biology*, 154(4):421–454, 1992a.

J. M. Savinell and B. Ø. Palsson. Network analysis of intermediary metabolism using linear optimization: II. interpretation of hybridoma cell metabolism. *Journal of Theoretical Biology*, 154(4):455–473, 1992b.

C. H. Schilling, D. Letscher, and B. Ø. Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology*, 203(3):229–248, 2000.

S. Schuster and D. Fell. *Bioinformatics—From Genomes to Therapies.*, volume 2, chapter 20, pages 755–805. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 3 volume set edition, 2007.

S. Schuster and C. Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(2):165–182, 1994.

S. Schuster and T. Höfer. Determining all extreme semi-positive conservation relations in chemical reaction systems: a test criterion for conservativity. *Journal of the Chemical Society, Faraday Transactions*, 87(16):2561–2566, 1991.

S. Schuster, T. Dandekar, and D. A. Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology*, 17(2):53–60, 1999.

S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18(3):326–332, 2000a.

S. Schuster, T. Pfeiffer, F. Moldenhauer, I. Koch, and T. Dandekar. Structural analysis of metabolic networks: elementary flux modes, analogy to Petri nets, and application to *Mycoplasma Pneumoniae*. In E. Bornberg-Bauer, U. Rost, and J. Stoye, editors, *Proceedings of the German Conference on Bioinformatics (GCB 2000), Heidelberg, October 5-7, 2000*, pages 115–120. Logos Verlag, 2000b. ISBN 3-89722-498-4.

S. Schuster, C. Hilgetag, J. H. Woods, and D. A. Fell. Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *Journal of Mathematical Biology*, 45(2):153–181, 2002.

R. Schwarz, C. Liang, C. Kaleta, M. Kühnel, E. Hoffmann, S. Kuznetsov, M. Hecker, G. Griffiths, S. Schuster, and T. Dandekar. Integrated network reconstruction, visualization and analysis using YANAsquare. *BMC Bioinformatics*, 8:313, 2007.

J. Schwender, F. Goffman, J. B. Ohlrogge, and Y. Shachar-Hill. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature*, 432(7018):779–782, 2004.

C. R. Scriver, A. L. Beaudet, and D. Valle. *The Metabolic and Molecular Bases of Inherited Disease.*, volume III. McGraw-Hill, New York, 1995.

D. Segrè, D. Vitkup, and G. M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(23):15112–15117, 2002.

P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11):2498–2504, 2003.

G. Shinar and M. Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–1391, 2010.

B. M. Slepchenko, J. C. Schaff, I. Macara, and L. M. Loew. Quantitative cell biology with the Virtual Cell. *Trends in Cell Biology*, 13(11):570–576, 2003.

A. G. Smart, L. A. N. Amaral, and J. M. Ottino. Cascading failure and robustness in metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36):13223–13228, 2008.

J. L. Snoep, F. Bruggeman, B. G. Olivier, and H. V. Westerhoff. Towards building the silicon cell: a modular approach. *Biosystems*, 83(2-3):207–216, 2006.

O. S. Soyer and T. Pfeiffer. Evolution under fluctuating environments explains observed robustness in metabolic networks. *PLoS Computational Biology*, 6(8):e1000907, 2010.

P. H. Starke. *Analyse von Petri-Netz-Modellen.* Teubner Verlag, Germany, 1990.

J. Stelling, S. Klamt, K. Bettenbrock, S. Schuster, and E. D. Gilles. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912):190–193, 2002.

J. Stelling, E. D. Gilles, and F. J. Doyle III. Robustness properties of circadian clock architectures. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13210–13215, 2004a.

J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle III, and J. Doyle. Robustness of cellular functions. *Cell*, 118(6):675–685, 2004b.

A.-M. Taburet, G. Raguin, C. Le Tiec, C. Droz, A. Barrail, I. Vincent, L. Morand-Joubert, G. Chêne, F. Clavel, and P.-M. Girard. Interactions between amprenavir and the lopinavir-ritonavir combination in heavily pretreated patients infected with human immunodeficiency virus. *Clinical Pharmacology & Therapeutics*, 75(4):310–323, 2004.

K. Takahashi, N. Ishikawa, Y. Sadamoto, H. Sasamoto, S. Ohta, A. Shiozawa, F. Miyoshi, Y. Naito, Y. Nakayama, and M. Tomita. E-Cell 2: multi-platform E-Cell simulation system. *Bioinformatics*, 19(13):1727–1729, 2003.

M. Terzer and J. Stelling. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24(19):2229–2235, 2008.

M. Tomita. Whole-cell simulation: a grand challenge of the 21st century. *Trends in Biotechnology*, 19(6):205–210, 2001.

C. T. Trinh and F. Srienc. Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Applied and Environmental Microbiology*, 75(21):6696–6705, 2009.

C. T. Trinh, P. Unrean, and F. Srienc. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Applied and Environmental Microbiology*, 74(12):3634–3643, 2008.

C. T. Trinh, A. Wlaschin, and F. Srienc. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Applied Microbiology and Biotechnology*, 81(5):813–826, 2009.

R. Urbanczik. SNA–a toolbox for the stoichiometric analysis of metabolic networks. *BMC Bioinformatics*, 7:129, 2006.

R. Urbanczik and C. Wagner. An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics*, 21(7):1203–1210, 2005.

N. A. W. van Riel. Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Briefings in Bioinformatics*, 7(4):364–374, 2006.

N. Vijayasankaran, R. Carlson, and F. Srienc. Metabolic pathway structures for recombinant protein synthesis in *Escherichia coli*. *Applied Microbiology and Biotechnology*, 68(6):737–746, 2005.

W. F. Visser, C. W. T. van Roermund, L. Ijlst, H. R. Waterham, and R. J. A. Wanders. Metabolite transport across the peroxisomal membrane. *Biochemical Journal*, 401(2):365–375, 2007.

D. Voet and J. G. Voet. *Biochemistry.* Wiley, New Jersey, 3rd edition, 2004.

A. von Kamp and S. Schuster. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22(15):1930–1931, 2006.

K. Voss, M. Heiner, and I. Koch. Steady state analysis of metabolic pathways using Petri nets. *In Silico Biology*, 3(3):367–387, 2003.

A. Wagner. *Robustness and Evolvability in Living Systems: (Princeton Studies in Complexity).* Princeton University Press, Princeton, New Jersey, 2005.

A. Werner and R. Heinrich. A kinetic model for the interaction of energy metabolism and osmotic states of human erythrocytes. Analysis of the stationary "in vivo" state and of time dependent variations under blood preservation conditions. *Biomedica biochimica acta*, 44(2):185–212, 1985.

H. V. Westerhoff and B. Ø. Palsson. The evolution of molecular biology into systems biology. *Nature Biotechnology*, 22(10):1249–1252, 2004.

T. Wilhelm, J. Behre, and S. Schuster. Analysis of structural robustness of metabolic networks. *IEE Proceedings - Systems Biology*, 1(1):114–120, 2004.

D. J. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature Reviews Genetics*, 10(2):122–133, 2009.

D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.-A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. Macinnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel, and L. Querengesser. HMDB: the Human Metabolome Database. *Nucleic Acids Research*, 35(Database issue):D521–D526, 2007.

J. Wolf, S. Becker-Weimann, and R. Heinrich. Analysing the robustness of cellular rhythms. *IEE Proceedings - Systems Biology*, 2(1):35–41, 2005.

J. Wu and E. Voit. Hybrid modeling in biochemical systems theory by means of functional Petri nets. *Journal of Bioinformatics and Computational Biology*, 7(1):107–134, 2009.

M. Xiong, J. Zhao, and H. Xiong. Network-based regulatory pathways analysis. *Bioinformatics*, 20(13):2056–2066, 2004.

I. Zevedei-Oancea and S. Schuster. Topological analysis of metabolic networks based on Petri net theory. *In Silico Biology*, 3(3):323–345, 2003.

Appendix

The Appendix contains the supplementary materials to Chapter 3 and to Chapter 4.

A Supplementary Material to Chapter 3

A.1 Supplement 1

Input file for the program `Metatool` of amino acid anabolism in *Escherichia coli*.
Identifiers: ENZREV, reversible enzymes; ENZIRREV, irreversible enzymes; METINT, internal metabolites; METEXT, external metabolites; CAT, catalysed reactions.

```
-ENZREV
R00114 R00214 R00228 R00230 R00236 R00245 R00248 R00258 R00268 R00341 R00354 R00355 R00371
R00405 R00472 R00586 R00667 R00694 R00707 R00734 R00945 R00999 R01061 R01071 R01073 R01082
R01090 R01215 R01248 R01323 R01324 R01512 R01513 R01698 R01715 R01777 R01899 R02164 R02283
R02413 R02570 R02735 R03084 R03243 R03313 R03443 R03460 R03968 R04001 R04173 R04440 R04475
R05068 R05069 R05071 spont2

-ENZIRREV
Ala_ex Arg_ex Asn_ex Asp_ex Cys_ex Ery_up G3P_up Gln_ex Glu_ex Gly_ex His_ex Ile_ex Leu_ex
Lys_ex Met_ex Phe_ex Pro_ex Pyr_up R00014 R00150 R00199 R00209 R00212 R00235 R00239 R00253
R00256 R00259 R00315 R00316 R00342 R00351 R00451 R00479 R00480 R00483 R00485 R00575 R00578
R00582 R00621 R00669 R00691 R00751 R00782 R00896 R00897 R00946 R00985 R00986 R00996 R01049
R01086 R01163 R01213 R01214 R01257 R01286 R01373 R01398 R01465 R01466 R01714 R01728 R01731
R01771 R01773 R01826 R01954 R02199 R02291 R02292 R02412 R02649 R02722 R02734 R03012 R03013
R03083 R03105 R03145 R03260 R03316 R03425 R03457 R03508 R03509 R03815 R04035 R04037 R04125
R04198 R04365 R04405 R04426 R04441 R04558 R04640 R04672 R04673 R05070 Rib_up Ser_ex SufS
Thr_ex Trp_ex Tyr_ex Val_ex

-METINT
```

```
(2R,3S)-3-Isopropylmalate (2S)-2-Isopropylmalate (3S)-Citryl-CoA
(R)-2,3-Dihydroxy-3-methylbutanoate (R)-2,3-Dihydroxy-3-methylpentanoate
(R)-3-Hydroxy-3-methyl-2-oxopentanoate (S)-1-Pyrroline-5-carboxylate
(S)-2-Aceto-2-hydroxybutanoate (S)-2-Acetolactate (S)-3-Methyl-2-oxopentanoic_acid
(S)-Malate 2,3,4,5-Tetrahydrodipicolinate 2-Hydroxyethyl-ThPP 2-Isopropylmaleate
2-Oxobutanoate 2-Oxoglutarate 3-(4-Hydroxyphenyl)pyruvate 3-Carboxy-1-hydroxypropyl-ThPP
3-Dehydroquinate 3-Dehydroshikimate 3-Hydroxy-3-methyl-2-oxobutanoic_acid
3-Methyl-2-oxobutanoic_acid 3-Phospho-D-glycerate 3-Phosphonooxypyruvate
4-Methyl-2-oxopentanoate 4-Phospho-L-aspartate Acetaldehyde Acetate Acetyl-CoA Acetyl-P
Acetyl_adenylate Anthranilate C01242 C01269 C01302 C04421 C04462 C04691 C04896 C04916
Carbamoyl-P Chorismate Citrate D-Erythrose-4P D-Ribose-5P Dihydrolipoamide
Dihydrolipoylprotein Fumarate Glyceraldehyde-3P Glycerate-1,3P2 Glycine Glyoxylate
Imidazole-acetol-P Imidazole-glycerol-3P Indoleglycerol-P Isocitrate
L-2,3-Dihydrodipicolinate L-2-Amino-3-oxobutanoic_acid L-Alanine L-Arginine L-Arogenate
L-Asparagine L-Aspartate L-Aspartate_4-semialdehyde L-Citrulline L-Cystathionine
L-Cysteine L-Glutamate L-Glutamate_5-semialdehyde L-Glutamine L-Glutamyl-P L-Histidinal
L-Histidine L-Histidinol L-Histidinol-P L-Homocysteine L-Homoserine L-Isoleucine L-Leucine
L-Lysine L-Methionine L-Ornithine L-Phenylalanine L-Proline L-Serine L-Threonine
L-Tryptophan L-Tyrosine L-Valine LL-2,6-Diaminoheptanedioate Lipoamide Lipoylprotein
Mercaptopyruvate N-(5-Phospho-D-ribosyl)anthranilate N-(L-Arginino)succinate
N-Acetyl-L-glutamate N-Acetyl-L-glutamate_5-semialdehyde N-Acetyl-L-glutamyl-P
N-Acetylornithine O-Acetyl-L-serine O-Phospho-L-homoserine O-Phospho-L-serine
O-Succinyl-L-homoserine Oxaloacetate Oxalosuccinate PRPP Phenylpyruvate
Phosphoenolpyruvate Phosphoribosyl-AMP Phosphoribosyl-ATP Prephenate Pyruvate
S-Succinyldihydrolipoamide Shikimate Shikimate-3P Succinate Succinyl-CoA ThPP
meso-2,6-Diaminoheptanedioate

-METEXT
5,10-Methylene-THF 5-Methyl-THF ADP AICAR AMP ATP C04144 C04489 CO2 CoA D-Erythrose-4P_ext
D-Ribose-5P_ext FAD FADH2 Ferricytochrome_ox Ferricytochrome_red Formate
Glyceraldehyde-3P_ext Glycine_ext H+ H2O H2S H2S2O3 L-Alanine_ext L-Arginine_ext
L-Asparagine_ext L-Aspartate_ext L-Cysteine_ext L-Glutamate_ext L-Glutamine_ext
L-Histidine_ext L-Isoleucine_ext L-Leucine_ext L-Lysine_ext L-Methionine_ext
L-Phenylalanine_ext L-Proline_ext L-Serine_ext L-Threonine_ext L-Tryptophan_ext
L-Tyrosine_ext L-Valine_ext NAD+ NADH NADP+ NADPH NH3 PPi Pi Pyruvate_ext Sulfite THF
Ubiquinol Ubiquinone [enzyme]-S-sulfanylcysteine [enzyme]-cysteine

-CAT
Ala_ex : L-Alanine = L-Alanine_ext .
Arg_ex : L-Arginine = L-Arginine_ext .
Asn_ex : L-Asparagine = L-Asparagine_ext .
Asp_ex : L-Aspartate = L-Aspartate_ext .
Cys_ex : L-Cysteine = L-Cysteine_ext .
Ery_up : D-Erythrose-4P_ext = D-Erythrose-4P .
G3P_up : Glyceraldehyde-3P_ext = Glyceraldehyde-3P .
Gln_ex : L-Glutamine = L-Glutamine_ext .
Glu_ex : L-Glutamate = L-Glutamate_ext .
Gly_ex : Glycine = Glycine_ext .
His_ex : L-Histidine = L-Histidine_ext .
Ile_ex : L-Isoleucine = L-Isoleucine_ext .
Leu_ex : L-Leucine = L-Leucine_ext .
```

```
Lys_ex : L-Lysine = L-Lysine_ext .
Met_ex : L-Methionine = L-Methionine_ext .
Phe_ex : L-Phenylalanine = L-Phenylalanine_ext .
Pro_ex : L-Proline = L-Proline_ext .
Pyr_up : Pyruvate_ext = Pyruvate .
R00014 : ThPP + Pyruvate = 2-Hydroxyethyl-ThPP + CO2 .
R00114 : 2 L-Glutamate + NADP+ = L-Glutamine + 2-Oxoglutarate + NADPH + H+ .
R00150 : ATP + NH3 + CO2 = ADP + Carbamoyl-P .
R00199 : ATP + Pyruvate + H2O = AMP + Phosphoenolpyruvate + Pi .
R00209 : Pyruvate + CoA + NAD+ = Acetyl-CoA + CO2 + NADH .
R00212 : CoA + Pyruvate = Acetyl-CoA + Formate .
R00214 : (S)-Malate + NAD+ = Pyruvate + CO2 + NADH .
R00228 : Acetaldehyde + CoA + NAD+ = Acetyl-CoA + NADH + H+ .
R00230 : Acetyl-CoA + Pi = CoA + Acetyl-P .
R00235 : ATP + Acetate + CoA = AMP + PPi + Acetyl-CoA .
R00236 : Acetyl_adenylate + CoA = AMP + Acetyl-CoA .
R00239 : ATP + L-Glutamate = ADP + L-Glutamyl-P .
R00245 : L-Glutamate_5-semialdehyde + NAD+ + H2O = L-Glutamate + NADH + H+ .
R00248 : L-Glutamate + NADP+ + H2O = 2-Oxoglutarate + NH3 + NADPH + H+ .
R00253 : ATP + L-Glutamate + NH3 = ADP + Pi + L-Glutamine .
R00256 : L-Glutamine + H2O = L-Glutamate + NH3 .
R00258 : L-Alanine + 2-Oxoglutarate = Pyruvate + L-Glutamate .
R00259 : Acetyl-CoA + L-Glutamate = CoA + N-Acetyl-L-glutamate .
R00268 : Oxalosuccinate = 2-Oxoglutarate + CO2 .
R00315 : ATP + Acetate = ADP + Acetyl-P .
R00316 : ATP + Acetate = PPi + Acetyl_adenylate .
R00341 : ATP + Oxaloacetate = ADP + Phosphoenolpyruvate + CO2 .
R00342 : (S)-Malate + NAD+ = Oxaloacetate + NADH + H+ .
R00351 : Acetyl-CoA + H2O + Oxaloacetate = Citrate + CoA .
R00354 : (3S)-Citryl-CoA = Acetyl-CoA + Oxaloacetate .
R00355 : L-Aspartate + 2-Oxoglutarate = Oxaloacetate + L-Glutamate .
R00371 : Acetyl-CoA + Glycine = CoA + L-2-Amino-3-oxobutanoic_acid .
R00405 : ATP + Succinate + CoA = ADP + Pi + Succinyl-CoA .
R00451 : meso-2,6-Diaminoheptanedioate = L-Lysine + CO2 .
R00472 : (S)-Malate + CoA = Acetyl-CoA + H2O + Glyoxylate .
R00479 : Isocitrate = Succinate + Glyoxylate .
R00480 : ATP + L-Aspartate = ADP + 4-Phospho-L-aspartate .
R00483 : ATP + L-Aspartate + NH3 = AMP + PPi + L-Asparagine .
R00485 : L-Asparagine + H2O = L-Aspartate + NH3 .
R00575 : 2 ATP + L-Glutamine + CO2 + 2 H2O = 2 ADP + Pi + L-Glutamate + Carbamoyl-P .
R00578 : ATP + L-Aspartate + L-Glutamine + H2O = AMP + PPi + L-Asparagine + L-Glutamate .
R00582 : O-Phospho-L-serine + H2O = L-Serine + Pi .
R00586 : L-Serine + Acetyl-CoA = O-Acetyl-L-serine + CoA .
R00621 : 2-Oxoglutarate + ThPP = 3-Carboxy-1-hydroxypropyl-ThPP + CO2 .
R00667 : L-Ornithine + 2-Oxoglutarate = L-Glutamate_5-semialdehyde + L-Glutamate .
R00669 : N-Acetylornithine + H2O = Acetate + L-Ornithine .
R00691 : L-Arogenate = L-Phenylalanine + H2O + CO2 .
R00694 : L-Phenylalanine + 2-Oxoglutarate = Phenylpyruvate + L-Glutamate .
R00707 : (S)-1-Pyrroline-5-carboxylate + NAD+ + 2 H2O = L-Glutamate + NADH + H+ .
R00734 : L-Tyrosine + 2-Oxoglutarate = 3-(4-Hydroxyphenyl)pyruvate + L-Glutamate .
R00751 : L-Threonine = Glycine + Acetaldehyde .
```

```
R00782 : H2S + Pyruvate + NH3 = L-Cysteine + H2O .
R00896 : Mercaptopyruvate + L-Glutamate = L-Cysteine + 2-Oxoglutarate .
R00897 : O-Acetyl-L-serine + H2S = L-Cysteine + Acetate .
R00945 : 5,10-Methylene-THF + Glycine + H2O = THF + L-Serine .
R00946 : 5-Methyl-THF + L-Homocysteine = THF + L-Methionine .
R00985 : Chorismate + NH3 = Anthranilate + Pyruvate + H2O .
R00986 : Chorismate + L-Glutamine = Anthranilate + Pyruvate + L-Glutamate .
R00996 : L-Threonine = 2-Oxobutanoate + NH3 .
R00999 : O-Succinyl-L-homoserine + H2O = 2-Oxobutanoate + Succinate + NH3 .
R01049 : ATP + D-Ribose-5P = AMP + PRPP .
R01061 : Glyceraldehyde-3P + Pi + NAD+ = Glycerate-1,3P2 + NADH + H+ .
R01071 : Phosphoribosyl-ATP + PPi = ATP + PRPP .
R01073 : N-(5-Phospho-D-ribosyl)anthranilate + PPi = Anthranilate + PRPP .
R01082 : (S)-Malate = Fumarate + H2O .
R01086 : N-(L-Arginino)succinate = Fumarate + L-Arginine .
R01090 : L-Leucine + 2-Oxoglutarate = 4-Methyl-2-oxopentanoate + L-Glutamate .
R01163 : L-Histidinal + H2O + 2 NAD+ = L-Histidine + 2 NADH + 2 H+ .
R01213 : Acetyl-CoA + 3-Methyl-2-oxobutanoic_acid + H2O = (2S)-2-Isopropylmalate + CoA .
R01214 : 3-Methyl-2-oxobutanoic_acid + L-Glutamate = L-Valine + 2-Oxoglutarate .
R01215 : L-Valine + Pyruvate = 3-Methyl-2-oxobutanoic_acid + L-Alanine .
R01248 : L-Proline + NAD+ = (S)-1-Pyrroline-5-carboxylate + NADH + H+ .
R01257 : (S)-Malate + FAD = FADH2 + Oxaloacetate .
R01286 : L-Cystathionine + H2O = L-Homocysteine + NH3 + Pyruvate .
R01323 : Acetyl-CoA + Citrate = Acetate + (3S)-Citryl-CoA .
R01324 : Citrate = Isocitrate .
R01373 : Prephenate = Phenylpyruvate + H2O + CO2 .
R01398 : Carbamoyl-P + L-Ornithine = Pi + L-Citrulline .
R01465 : L-Threonine + NAD+ = L-2-Amino-3-oxobutanoic_acid + NADH + H+ .
R01466 : O-Phospho-L-homoserine + H2O = L-Threonine + Pi .
R01512 : ATP + 3-Phospho-D-glycerate = ADP + Glycerate-1,3P2 .
R01513 : 3-Phospho-D-glycerate + NAD+ = 3-Phosphonooxypyruvate + NADH + H+ .
R01698 : Dihydrolipoamide + NAD+ = Lipoamide + NADH + H+ .
R01714 : C01269 = Chorismate + Pi .
R01715 : Chorismate = Prephenate .
R01728 : Prephenate + NAD+ = 3-(4-Hydroxyphenyl)pyruvate + CO2 + NADH + H+ .
R01731 : L-Aspartate + Prephenate = Oxaloacetate + L-Arogenate .
R01771 : ATP + L-Homoserine = ADP + O-Phospho-L-homoserine .
R01773 : L-Aspartate_4-semialdehyde + NADH + H+ = L-Homoserine + NAD+ .
R01777 : Succinyl-CoA + L-Homoserine = CoA + O-Succinyl-L-homoserine .
R01826 : Phosphoenolpyruvate + D-Erythrose-4P + H2O = C04691 + Pi .
R01899 : Isocitrate + NADP+ = Oxalosuccinate + NADPH + H+ .
R01954 : ATP + L-Citrulline + L-Aspartate = AMP + PPi + N-(L-Arginino)succinate .
R02164 : Ubiquinone + Succinate = Ubiquinol + Fumarate .
R02199 : (S)-3-Methyl-2-oxopentanoic_acid + L-Glutamate = L-Isoleucine + 2-Oxoglutarate .
R02283 : N-Acetylornithine + 2-Oxoglutarate = N-Acetyl-L-glutamate_5-semialdehyde +
         L-Glutamate .
R02291 : 4-Phospho-L-aspartate + NADPH + H+ = L-Aspartate_4-semialdehyde + Pi + NADP+ .
R02292 : L-Aspartate_4-semialdehyde + Pyruvate = L-2,3-Dihydrodipicolinate + 2 H2O .
R02412 : ATP + Shikimate = ADP + Shikimate-3P .
R02413 : Shikimate + NADP+ = 3-Dehydroshikimate + NADPH + H+ .
R02570 : Succinyl-CoA + Dihydrolipoamide = CoA + S-Succinyldihydrolipoamide .
```

```
R02649 : ATP + N-Acetyl-L-glutamate = ADP + N-Acetyl-L-glutamyl-P .
R02722 : L-Serine + Indoleglycerol-P = L-Tryptophan + Glyceraldehyde-3P + H2O .
R02734 : C04421 + H2O = Succinate + LL-2,6-Diaminoheptanedioate .
R02735 : LL-2,6-Diaminoheptanedioate = meso-2,6-Diaminoheptanedioate .
R03012 : L-Histidinol + NAD+ = L-Histidinal + NADH + H+ .
R03013 : L-Histidinol-P + H2O = L-Histidinol + Pi .
R03083 : C04691 = 3-Dehydroquinate + Pi .
R03084 : 3-Dehydroquinate = 3-Dehydroshikimate + H2O .
R03105 : H2S2O3 + Pyruvate = Mercaptopyruvate + Sulfite .
R03145 : Ferricytochrome_ox + Pyruvate + H2O = Ferricytochrome_red + Acetate + CO2 .
R03243 : L-Histidinol-P + 2-Oxoglutarate = Imidazole-acetol-P + L-Glutamate .
R03260 : O-Succinyl-L-homoserine + L-Cysteine = L-Cystathionine + Succinate .
R03313 : L-Glutamate_5-semialdehyde + Pi + NADP+ = L-Glutamyl-P + NADPH + H+ .
R03316 : 3-Carboxy-1-hydroxypropyl-ThPP + Lipoamide = S-Succinyldihydrolipoamide + ThPP .
R03425 : Glycine + Lipoylprotein = C01242 + CO2 .
R03443 : N-Acetyl-L-glutamate_5-semialdehyde + Pi + NADP+ = N-Acetyl-L-glutamyl-P +
         NADPH + H+ .
R03457 : Imidazole-glycerol-3P = Imidazole-acetol-P + H2O .
R03460 : Phosphoenolpyruvate + Shikimate-3P = Pi + C01269 .
R03508 : C01302 = Indoleglycerol-P + CO2 + H2O .
R03509 : N-(5-Phospho-D-ribosyl)anthranilate = C01302 .
R03815 : Dihydrolipoylprotein + NAD+ = Lipoylprotein + NADH + H+ .
R03968 : (2S)-2-Isopropylmalate = 2-Isopropylmaleate + H2O .
R04001 : (2R,3S)-3-Isopropylmalate = 2-Isopropylmaleate + H2O .
R04035 : Phosphoribosyl-ATP + H2O = Phosphoribosyl-AMP + PPi .
R04037 : Phosphoribosyl-AMP + H2O = C04896 .
R04125 : THF + C01242 = 5,10-Methylene-THF + NH3 + Dihydrolipoylprotein .
R04173 : O-Phospho-L-serine + 2-Oxoglutarate = 3-Phosphonooxypyruvate + L-Glutamate .
R04198 : L-2,3-Dihydrodipicolinate + NADH + H+ = 2,3,4,5-Tetrahydrodipicolinate + NAD+ .
R04365 : Succinyl-CoA + 2,3,4,5-Tetrahydrodipicolinate + H2O = CoA + C04462 .
R04405 : C04489 + L-Homocysteine = C04144 + L-Methionine .
R04426 : (2R,3S)-3-Isopropylmalate + NAD+ = 4-Methyl-2-oxopentanoate + NADH + H+ + CO2 .
R04440 : (R)-2,3-Dihydroxy-3-methylbutanoate + NADP+ =
         3-Hydroxy-3-methyl-2-oxobutanoic_acid + NADPH .
R04441 : (R)-2,3-Dihydroxy-3-methylbutanoate = 3-Methyl-2-oxobutanoic_acid + H2O .
R04475 : C04421 + 2-Oxoglutarate = C04462 + L-Glutamate .
R04558 : C04916 + L-Glutamine = AICAR + L-Glutamate + Imidazole-glycerol-3P .
R04640 : C04896 = C04916 .
R04672 : 2-Hydroxyethyl-ThPP + Pyruvate = (S)-2-Acetolactate + ThPP .
R04673 : 2-Oxobutanoate + 2-Hydroxyethyl-ThPP = (S)-2-Aceto-2-hydroxybutanoate + ThPP .
R05068 : (R)-2,3-Dihydroxy-3-methylpentanoate + NADP+ =
         (R)-3-Hydroxy-3-methyl-2-oxopentanoate + NADPH + H+ .
R05069 : (S)-2-Aceto-2-hydroxybutanoate = (R)-3-Hydroxy-3-methyl-2-oxopentanoate .
R05070 : (R)-2,3-Dihydroxy-3-methylpentanoate = (S)-3-Methyl-2-oxopentanoic_acid + H2O .
R05071 : (S)-2-Acetolactate = 3-Hydroxy-3-methyl-2-oxobutanoic_acid .
Rib_up : D-Ribose-5P_ext = D-Ribose-5P .
Ser_ex : L-Serine = L-Serine_ext .
SufS : L-Cysteine + [enzyme]-cysteine = L-Alanine + [enzyme]-S-sulfanylcysteine .
Thr_ex : L-Threonine = L-Threonine_ext .
Trp_ex : L-Tryptophan = L-Tryptophan_ext .
Tyr_ex : L-Tyrosine = L-Tyrosine_ext .
```

```
Val_ex : L-Valine = L-Valine_ext .
spont2 : L-Glutamate_5-semialdehyde = (S)-1-Pyrroline-5-carboxylate + H2O .
```

A.2 Supplement 2

Input file for the program `Metatool` of the *Escherichia coli* subsystem comprising those amino acids which are essential for humans.
Identifiers have the same meaning as given in Supplement 1.
For detailed descriptions refer to main text.

```
-ENZREV
R00214 R00230 R00236 R00248 R00268 R00341 R00354 R00355 R00405 R00472 R00694 R00734 R00999
R01071 R01073 R01082 R01090 R01323 R01324 R01698 R01715 R01777 R01899 R02164 R02413 R02570
R02735 R03084 R03243 R03460 R03968 R04001 R04440 R04475 R05068 R05069 R05071

-ENZIRREV
Cys_up Ery_up His_ex Ile_ex Leu_ex Lys_ex Met_ex Phe_ex Pyr_up R00014 R00199 R00209 R00212
R00235 R00253 R00315 R00316 R00342 R00351 R00451 R00479 R00480 R00621 R00691 R00946 R00985
R00986 R01049 R01163 R01213 R01214 R01257 R01286 R01373 R01466 R01714 R01728 R01731 R01771
R01773 R01826 R02199 R02291 R02292 R02412 R02722 R02734 R03012 R03013 R03083 R03145 R03260
R03316 R03457 R03508 R03509 R04035 R04037 R04198 R04365 R04405 R04426 R04441 R04558 R04640
R04672 R04673 R05070 Rib_up Ser_up Thr_ex Trp_ex Tyr_ex Val_ex

-METINT
(2R,3S)-3-Isopropylmalate (2S)-2-Isopropylmalate (3S)-Citryl-CoA
(R)-2,3-Dihydroxy-3-methylbutanoate (R)-2,3-Dihydroxy-3-methylpentanoate
(R)-3-Hydroxy-3-methyl-2-oxopentanoate (S)-2-Aceto-2-hydroxybutanoate (S)-2-Acetolactate
(S)-3-Methyl-2-oxopentanoic_acid (S)-Malate 2,3,4,5-Tetrahydrodipicolinate
2-Hydroxyethyl-ThPP 2-Isopropylmaleate 2-Oxobutanoate 2-Oxoglutarate
3-(4-Hydroxyphenyl)pyruvate 3-Carboxy-1-hydroxypropyl-ThPP 3-Dehydroquinate
3-Dehydroshikimate 3-Hydroxy-3-methyl-2-oxobutanoic_acid 3-Methyl-2-oxobutanoic_acid
4-Methyl-2-oxopentanoate 4-Phospho-L-aspartate Acetate Acetyl-CoA Acetyl-P
Acetyl_adenylate Anthranilate C01269 C01302 C04421 C04462 C04691 C04896 C04916 Chorismate
Citrate D-Erythrose-4P D-Ribose-5P Dihydrolipoamide Fumarate Glyoxylate Imidazole-acetol-P
Imidazole-glycerol-3P Indoleglycerol-P Isocitrate L-2,3-Dihydrodipicolinate L-Arogenate
L-Aspartate_4-semialdehyde L-Cystathionine L-Cysteine L-Glutamine L-Histidinal L-Histidine
L-Histidinol L-Histidinol-P L-Homocysteine L-Homoserine L-Isoleucine L-Leucine L-Lysine
L-Methionine L-Phenylalanine L-Serine L-Threonine L-Tryptophan L-Tyrosine L-Valine
LL-2,6-Diaminoheptanedioate Lipoamide N-(5-Phospho-D-ribosyl)anthranilate
O-Phospho-L-homoserine O-Succinyl-L-homoserine Oxaloacetate Oxalosuccinate PRPP
Phenylpyruvate Phosphoenolpyruvate Phosphoribosyl-AMP Phosphoribosyl-ATP Prephenate
Pyruvate S-Succinyldihydrolipoamide Shikimate Shikimate-3P Succinate Succinyl-CoA ThPP
meso-2,6-Diaminoheptanedioate

-METEXT
5-Methyl-THF ADP AICAR AMP ATP C04144 C04489 CO2 CoA D-Erythrose-4P_ext D-Ribose-5P_ext
```

```
FAD FADH2 Ferricytochrome_ox Ferricytochrome_red Formate Glyceraldehyde-3P H+ H2O
L-Aspartate L-Cysteine_ext L-Glutamate L-Histidine_ext L-Isoleucine_ext L-Leucine_ext
L-Lysine_ext L-Methionine_ext L-Phenylalanine_ext L-Serine_ext L-Threonine_ext
L-Tryptophan_ext L-Tyrosine_ext L-Valine_ext NAD+ NADH NADP+ NADPH NH3 PPi Pi Pyruvate_ext
THF Ubiquinol Ubiquinone

-CAT
Cys_up : L-Cysteine_ext = L-Cysteine .
Ery_up : D-Erythrose-4P_ext = D-Erythrose-4P .
His_ex : L-Histidine = L-Histidine_ext .
Ile_ex : L-Isoleucine = L-Isoleucine_ext .
Leu_ex : L-Leucine = L-Leucine_ext .
Lys_ex : L-Lysine = L-Lysine_ext .
Met_ex : L-Methionine = L-Methionine_ext .
Phe_ex : L-Phenylalanine = L-Phenylalanine_ext .
Pyr_up : Pyruvate_ext = Pyruvate .
R00014 : ThPP + Pyruvate = 2-Hydroxyethyl-ThPP + CO2 .
R00199 : ATP + Pyruvate + H2O = AMP + Phosphoenolpyruvate + Pi .
R00209 : Pyruvate + CoA + NAD+ = Acetyl-CoA + CO2 + NADH .
R00212 : CoA + Pyruvate = Acetyl-CoA + Formate .
R00214 : (S)-Malate + NAD+ = Pyruvate + CO2 + NADH .
R00230 : Acetyl-CoA + Pi = CoA + Acetyl-P .
R00235 : ATP + Acetate + CoA = AMP + PPi + Acetyl-CoA .
R00236 : Acetyl_adenylate + CoA = AMP + Acetyl-CoA .
R00248 : L-Glutamate + NADP+ + H2O = 2-Oxoglutarate + NH3 + NADPH + H+ .
R00253 : ATP + L-Glutamate + NH3 = ADP + Pi + L-Glutamine .
R00268 : Oxalosuccinate = 2-Oxoglutarate + CO2 .
R00315 : ATP + Acetate = ADP + Acetyl-P .
R00316 : ATP + Acetate = PPi + Acetyl_adenylate .
R00341 : ATP + Oxaloacetate = ADP + Phosphoenolpyruvate + CO2 .
R00342 : (S)-Malate + NAD+ = Oxaloacetate + NADH + H+ .
R00351 : Acetyl-CoA + H2O + Oxaloacetate = Citrate + CoA .
R00354 : (3S)-Citryl-CoA = Acetyl-CoA + Oxaloacetate .
R00355 : L-Aspartate + 2-Oxoglutarate = Oxaloacetate + L-Glutamate .
R00405 : ATP + Succinate + CoA = ADP + Pi + Succinyl-CoA .
R00451 : meso-2,6-Diaminoheptanedioate = L-Lysine + CO2 .
R00472 : (S)-Malate + CoA = Acetyl-CoA + H2O + Glyoxylate .
R00479 : Isocitrate = Succinate + Glyoxylate .
R00480 : ATP + L-Aspartate = ADP + 4-Phospho-L-aspartate .
R00621 : 2-Oxoglutarate + ThPP = 3-Carboxy-1-hydroxypropyl-ThPP + CO2 .
R00691 : L-Arogenate = L-Phenylalanine + H2O + CO2 .
R00694 : L-Phenylalanine + 2-Oxoglutarate = Phenylpyruvate + L-Glutamate .
R00734 : L-Tyrosine + 2-Oxoglutarate = 3-(4-Hydroxyphenyl)pyruvate + L-Glutamate .
R00946 : 5-Methyl-THF + L-Homocysteine = THF + L-Methionine .
R00985 : Chorismate + NH3 = Anthranilate + Pyruvate + H2O .
R00986 : Chorismate + L-Glutamine = Anthranilate + Pyruvate + L-Glutamate .
R00999 : O-Succinyl-L-homoserine + H2O = 2-Oxobutanoate + Succinate + NH3 .
R01049 : ATP + D-Ribose-5P = AMP + PRPP .
R01071 : Phosphoribosyl-ATP + PPi = ATP + PRPP .
R01073 : N-(5-Phospho-D-ribosyl)anthranilate + PPi = Anthranilate + PRPP .
R01082 : (S)-Malate = Fumarate + H2O .
```

```
R01090 : L-Leucine + 2-Oxoglutarate = 4-Methyl-2-oxopentanoate + L-Glutamate .
R01163 : L-Histidinal + H2O + 2 NAD+ = L-Histidine + 2 NADH + 2 H+ .
R01213 : Acetyl-CoA + 3-Methyl-2-oxobutanoic_acid + H2O = (2S)-2-Isopropylmalate + CoA .
R01214 : 3-Methyl-2-oxobutanoic_acid + L-Glutamate = L-Valine + 2-Oxoglutarate .
R01257 : (S)-Malate + FAD = FADH2 + Oxaloacetate .
R01286 : L-Cystathionine + H2O = L-Homocysteine + NH3 + Pyruvate .
R01323 : Acetyl-CoA + Citrate = Acetate + (3S)-Citryl-CoA .
R01324 : Citrate = Isocitrate .
R01373 : Prephenate = Phenylpyruvate + H2O + CO2 .
R01466 : O-Phospho-L-homoserine + H2O = L-Threonine + Pi .
R01698 : Dihydrolipoamide + NAD+ = Lipoamide + NADH + H+ .
R01714 : C01269 = Chorismate + Pi .
R01715 : Chorismate = Prephenate .
R01728 : Prephenate + NAD+ = 3-(4-Hydroxyphenyl)pyruvate + CO2 + NADH + H+ .
R01731 : L-Aspartate + Prephenate = Oxaloacetate + L-Arogenate .
R01771 : ATP + L-Homoserine = ADP + O-Phospho-L-homoserine .
R01773 : L-Aspartate_4-semialdehyde + NADH + H+ = L-Homoserine + NAD+ .
R01777 : Succinyl-CoA + L-Homoserine = CoA + O-Succinyl-L-homoserine .
R01826 : Phosphoenolpyruvate + D-Erythrose-4P + H2O = C04691 + Pi .
R01899 : Isocitrate + NADP+ = Oxalosuccinate + NADPH + H+ .
R02164 : Ubiquinone + Succinate = Ubiquinol + Fumarate .
R02199 : (S)-3-Methyl-2-oxopentanoic_acid + L-Glutamate = L-Isoleucine + 2-Oxoglutarate .
R02291 : 4-Phospho-L-aspartate + NADPH + H+ = L-Aspartate_4-semialdehyde + Pi + NADP+ .
R02292 : L-Aspartate_4-semialdehyde + Pyruvate = L-2,3-Dihydrodipicolinate + 2 H2O .
R02412 : ATP + Shikimate = ADP + Shikimate-3P .
R02413 : Shikimate + NADP+ = 3-Dehydroshikimate + NADPH + H+ .
R02570 : Succinyl-CoA + Dihydrolipoamide = CoA + S-Succinyldihydrolipoamide .
R02722 : L-Serine + Indoleglycerol-P = L-Tryptophan + Glyceraldehyde-3P + H2O .
R02734 : C04421 + H2O = Succinate + LL-2,6-Diaminoheptanedioate .
R02735 : LL-2,6-Diaminoheptanedioate = meso-2,6-Diaminoheptanedioate .
R03012 : L-Histidinol + NAD+ = L-Histidinal + NADH + H+ .
R03013 : L-Histidinol-P + H2O = L-Histidinol + Pi .
R03083 : C04691 = 3-Dehydroquinate + Pi .
R03084 : 3-Dehydroquinate = 3-Dehydroshikimate + H2O .
R03145 : Ferricytochrome_ox + Pyruvate + H2O = Ferricytochrome_red + Acetate + CO2 .
R03243 : L-Histidinol-P + 2-Oxoglutarate = Imidazole-acetol-P + L-Glutamate .
R03260 : O-Succinyl-L-homoserine + L-Cysteine = L-Cystathionine + Succinate .
R03316 : 3-Carboxy-1-hydroxypropyl-ThPP + Lipoamide = S-Succinyldihydrolipoamide + ThPP .
R03457 : Imidazole-glycerol-3P = Imidazole-acetol-P + H2O .
R03460 : Phosphoenolpyruvate + Shikimate-3P = Pi + C01269 .
R03508 : C01302 = Indoleglycerol-P + CO2 + H2O .
R03509 : N-(5-Phospho-D-ribosyl)anthranilate = C01302 .
R03968 : (2S)-2-Isopropylmalate = 2-Isopropylmaleate + H2O .
R04001 : (2R,3S)-3-Isopropylmalate = 2-Isopropylmaleate + H2O .
R04035 : Phosphoribosyl-ATP + H2O = Phosphoribosyl-AMP + PPi .
R04037 : Phosphoribosyl-AMP + H2O = C04896 .
R04198 : L-2,3-Dihydrodipicolinate + NADH + H+ = 2,3,4,5-Tetrahydrodipicolinate + NAD+ .
R04365 : Succinyl-CoA + 2,3,4,5-Tetrahydrodipicolinate + H2O = CoA + C04462 .
R04405 : C04489 + L-Homocysteine = C04144 + L-Methionine .
R04426 : (2R,3S)-3-Isopropylmalate + NAD+ = 4-Methyl-2-oxopentanoate + NADH + H+ + CO2 .
R04440 : (R)-2,3-Dihydroxy-3-methylbutanoate + NADP+ =
```

```
        3-Hydroxy-3-methyl-2-oxobutanoic_acid + NADPH .
R04441 : (R)-2,3-Dihydroxy-3-methylbutanoate = 3-Methyl-2-oxobutanoic_acid + H2O .
R04475 : C04421 + 2-Oxoglutarate = C04462 + L-Glutamate .
R04558 : C04916 + L-Glutamine = AICAR + L-Glutamate + Imidazole-glycerol-3P .
R04640 : C04896 = C04916 .
R04672 : 2-Hydroxyethyl-ThPP + Pyruvate = (S)-2-Acetolactate + ThPP .
R04673 : 2-Oxobutanoate + 2-Hydroxyethyl-ThPP = (S)-2-Aceto-2-hydroxybutanoate + ThPP .
R05068 : (R)-2,3-Dihydroxy-3-methylpentanoate + NADP+ =
        (R)-3-Hydroxy-3-methyl-2-oxopentanoate + NADPH + H+ .
R05069 : (S)-2-Aceto-2-hydroxybutanoate = (R)-3-Hydroxy-3-methyl-2-oxopentanoate .
R05070 : (R)-2,3-Dihydroxy-3-methylpentanoate = (S)-3-Methyl-2-oxopentanoic_acid + H2O .
R05071 : (S)-2-Acetolactate = 3-Hydroxy-3-methyl-2-oxobutanoic_acid .
Rib_up : D-Ribose-5P_ext = D-Ribose-5P .
Ser_up : L-Serine_ext = L-Serine .
Thr_ex : L-Threonine = L-Threonine_ext .
Trp_ex : L-Tryptophan = L-Tryptophan_ext .
Tyr_ex : L-Tyrosine = L-Tyrosine_ext .
Val_ex : L-Valine = L-Valine_ext .
```

A.3 Supplement 3

Input file for the program `Metatool` of the *Escherichia coli* subsystem comprising those amino acids which are non-essential for humans.
Identifiers have the same meaning as given in Supplement 1.
For detailed descriptions refer to main text.

```
-ENZREV
R00114 R00214 R00230 R00236 R00245 R00248 R00258 R00268 R00341 R00354 R00355 R00405 R00472
R00586 R00667 R00707 R00945 R01061 R01082 R01248 R01323 R01324 R01512 R01513 R01698 R01899
R02164 R02283 R02570 R03313 R03443 R04173 spont2

-ENZIRREV
Ala_ex Arg_ex Asn_ex Asp_ex Cys_ex G3P_up Gln_ex Glu_ex Gly_ex Pro_ex Pyr_up R00150 R00199
R00209 R00212 R00235 R00239 R00253 R00256 R00259 R00315 R00316 R00342 R00351 R00479 R00483
R00485 R00575 R00578 R00582 R00621 R00669 R00782 R00896 R00897 R01086 R01257 R01398 R01954
R02649 R03105 R03145 R03316 R03425 R03815 R04125 Ser_ex SufS

-METINT
(3S)-Citryl-CoA (S)-1-Pyrroline-5-carboxylate (S)-Malate 2-Oxoglutarate
3-Carboxy-1-hydroxypropyl-ThPP 3-Phospho-D-glycerate 3-Phosphonooxypyruvate Acetate
Acetyl-CoA Acetyl-P Acetyl_adenylate C01242 Carbamoyl-P Citrate Dihydrolipoamide
Dihydrolipoylprotein Fumarate Glyceraldehyde-3P Glycerate-1,3P2 Glycine Glyoxylate
Isocitrate L-Alanine L-Arginine L-Asparagine L-Aspartate L-Citrulline L-Cysteine
L-Glutamate L-Glutamate_5-semialdehyde L-Glutamine L-Glutamyl-P L-Ornithine L-Proline
L-Serine Lipoamide Lipoylprotein Mercaptopyruvate N-(L-Arginino)succinate
N-Acetyl-L-glutamate N-Acetyl-L-glutamate_5-semialdehyde N-Acetyl-L-glutamyl-P
```

```
N-Acetylornithine O-Acetyl-L-serine O-Phospho-L-serine Oxaloacetate Oxalosuccinate
Phosphoenolpyruvate Pyruvate S-Succinyldihydrolipoamide Succinate Succinyl-CoA ThPP

-METEXT
5,10-Methylene-THF ADP AMP ATP CO2 CoA FAD FADH2 Ferricytochrome_ox Ferricytochrome_red
Formate Glyceraldehyde-3P_ext Glycine_ext H+ H2O H2S H2S2O3 L-Alanine_ext L-Arginine_ext
L-Asparagine_ext L-Aspartate_ext L-Cysteine_ext L-Glutamate_ext L-Glutamine_ext
L-Proline_ext L-Serine_ext NAD+ NADH NADP+ NADPH NH3 PPi Pi Pyruvate_ext Sulfite THF
Ubiquinol Ubiquinone [enzyme]-S-sulfanylcysteine [enzyme]-cysteine

-CAT
Ala_ex : L-Alanine = L-Alanine_ext .
Arg_ex : L-Arginine = L-Arginine_ext .
Asn_ex : L-Asparagine = L-Asparagine_ext .
Asp_ex : L-Aspartate = L-Aspartate_ext .
Cys_ex : L-Cysteine = L-Cysteine_ext .
G3P_up : Glyceraldehyde-3P_ext = Glyceraldehyde-3P .
Gln_ex : L-Glutamine = L-Glutamine_ext .
Glu_ex : L-Glutamate = L-Glutamate_ext .
Gly_ex : Glycine = Glycine_ext .
Pro_ex : L-Proline = L-Proline_ext .
Pyr_up : Pyruvate_ext = Pyruvate .
R00114 : 2 L-Glutamate + NADP+ = L-Glutamine + 2-Oxoglutarate + NADPH + H+ .
R00150 : ATP + NH3 + CO2 = ADP + Carbamoyl-P .
R00199 : ATP + Pyruvate + H2O = AMP + Phosphoenolpyruvate + Pi .
R00209 : Pyruvate + CoA + NAD+ = Acetyl-CoA + CO2 + NADH .
R00212 : CoA + Pyruvate = Acetyl-CoA + Formate .
R00214 : (S)-Malate + NAD+ = Pyruvate + CO2 + NADH .
R00230 : Acetyl-CoA + Pi = CoA + Acetyl-P .
R00235 : ATP + Acetate + CoA = AMP + PPi + Acetyl-CoA .
R00236 : Acetyl_adenylate + CoA = AMP + Acetyl-CoA .
R00239 : ATP + L-Glutamate = ADP + L-Glutamyl-P .
R00245 : L-Glutamate_5-semialdehyde + NAD+ + H2O = L-Glutamate + NADH + H+ .
R00248 : L-Glutamate + NADP+ + H2O = 2-Oxoglutarate + NH3 + NADPH + H+ .
R00253 : ATP + L-Glutamate + NH3 = ADP + Pi + L-Glutamine .
R00256 : L-Glutamine + H2O = L-Glutamate + NH3 .
R00258 : L-Alanine + 2-Oxoglutarate = Pyruvate + L-Glutamate .
R00259 : Acetyl-CoA + L-Glutamate = CoA + N-Acetyl-L-glutamate .
R00268 : Oxalosuccinate = 2-Oxoglutarate + CO2 .
R00315 : ATP + Acetate = ADP + Acetyl-P .
R00316 : ATP + Acetate = PPi + Acetyl_adenylate .
R00341 : ATP + Oxaloacetate = ADP + Phosphoenolpyruvate + CO2 .
R00342 : (S)-Malate + NAD+ = Oxaloacetate + NADH + H+ .
R00351 : Acetyl-CoA + H2O + Oxaloacetate = Citrate + CoA .
R00354 : (3S)-Citryl-CoA = Acetyl-CoA + Oxaloacetate .
R00355 : L-Aspartate + 2-Oxoglutarate = Oxaloacetate + L-Glutamate .
R00405 : ATP + Succinate + CoA = ADP + Pi + Succinyl-CoA .
R00472 : (S)-Malate + CoA = Acetyl-CoA + H2O + Glyoxylate .
R00479 : Isocitrate = Succinate + Glyoxylate .
R00483 : ATP + L-Aspartate + NH3 = AMP + PPi + L-Asparagine .
R00485 : L-Asparagine + H2O = L-Aspartate + NH3 .
```

```
R00575 : 2 ATP + L-Glutamine + CO2 + 2 H2O = 2 ADP + Pi + L-Glutamate + Carbamoyl-P .
R00578 : ATP + L-Aspartate + L-Glutamine + H2O = AMP + PPi + L-Asparagine + L-Glutamate .
R00582 : O-Phospho-L-serine + H2O = L-Serine + Pi .
R00586 : L-Serine + Acetyl-CoA = O-Acetyl-L-serine + CoA .
R00621 : 2-Oxoglutarate + ThPP = 3-Carboxy-1-hydroxypropyl-ThPP + CO2 .
R00667 : L-Ornithine + 2-Oxoglutarate = L-Glutamate_5-semialdehyde + L-Glutamate .
R00669 : N-Acetylornithine + H2O = Acetate + L-Ornithine .
R00707 : (S)-1-Pyrroline-5-carboxylate + NAD+ + 2 H2O = L-Glutamate + NADH + H+ .
R00782 : H2S + Pyruvate + NH3 = L-Cysteine + H2O .
R00896 : Mercaptopyruvate + L-Glutamate = L-Cysteine + 2-Oxoglutarate .
R00897 : O-Acetyl-L-serine + H2S = L-Cysteine + Acetate .
R00945 : 5,10-Methylene-THF + Glycine + H2O = THF + L-Serine .
R01061 : Glyceraldehyde-3P + Pi + NAD+ = Glycerate-1,3P2 + NADH + H+ .
R01082 : (S)-Malate = Fumarate + H2O .
R01086 : N-(L-Arginino)succinate = Fumarate + L-Arginine .
R01248 : L-Proline + NAD+ = (S)-1-Pyrroline-5-carboxylate + NADH + H+ .
R01257 : (S)-Malate + FAD = FADH2 + Oxaloacetate .
R01323 : Acetyl-CoA + Citrate = Acetate + (3S)-Citryl-CoA .
R01324 : Citrate = Isocitrate .
R01398 : Carbamoyl-P + L-Ornithine = Pi + L-Citrulline .
R01512 : ATP + 3-Phospho-D-glycerate = ADP + Glycerate-1,3P2 .
R01513 : 3-Phospho-D-glycerate + NAD+ = 3-Phosphonooxypyruvate + NADH + H+ .
R01698 : Dihydrolipoamide + NAD+ = Lipoamide + NADH + H+ .
R01899 : Isocitrate + NADP+ = Oxalosuccinate + NADPH + H+ .
R01954 : ATP + L-Citrulline + L-Aspartate = AMP + PPi + N-(L-Arginino)succinate .
R02164 : Ubiquinone + Succinate = Ubiquinol + Fumarate .
R02283 : N-Acetylornithine + 2-Oxoglutarate = N-Acetyl-L-glutamate_5-semialdehyde +
         L-Glutamate .
R02570 : Succinyl-CoA + Dihydrolipoamide = CoA + S-Succinyldihydrolipoamide .
R02649 : ATP + N-Acetyl-L-glutamate = ADP + N-Acetyl-L-glutamyl-P .
R03105 : H2S2O3 + Pyruvate = Mercaptopyruvate + Sulfite .
R03145 : Ferricytochrome_ox + Pyruvate + H2O = Ferricytochrome_red + Acetate + CO2 .
R03313 : L-Glutamate_5-semialdehyde + Pi + NADP+ = L-Glutamyl-P + NADPH + H+ .
R03316 : 3-Carboxy-1-hydroxypropyl-ThPP + Lipoamide = S-Succinyldihydrolipoamide + ThPP .
R03425 : Glycine + Lipoylprotein = C01242 + CO2 .
R03443 : N-Acetyl-L-glutamate_5-semialdehyde + Pi + NADP+ = N-Acetyl-L-glutamyl-P +
         NADPH + H+ .
R03815 : Dihydrolipoylprotein + NAD+ = Lipoylprotein + NADH + H+ .
R04125 : THF + C01242 = 5,10-Methylene-THF + NH3 + Dihydrolipoylprotein .
R04173 : O-Phospho-L-serine + 2-Oxoglutarate = 3-Phosphonooxypyruvate + L-Glutamate .
Ser_ex : L-Serine = L-Serine_ext .
SufS : L-Cysteine + [enzyme]-cysteine = L-Alanine + [enzyme]-S-sulfanylcysteine .
spont2 : L-Glutamate_5-semialdehyde = (S)-1-Pyrroline-5-carboxylate + H2O .
```

A.4 Supplement 4

Input file for the program `Metatool` of the human hepatocyte amino acid system.

Identifiers have the same meaning as given in Supplement 1.

```
-ENZREV
Ala_diff Cys_diff DIC_1 DIC_2 Gly_diff OGC ORC_2 Pro_diff Pro_tr R00214 R00243 R00245
R00258 R00268 R00344 R00355 R00405 R00430 R00432 R00667 R00707 R00709 R00945 R01061 R01082
R01248 R01324 R01512 R01513 R01698 R01899 R02164 R02570 R03313 R04173 Ser_diff spont2

-ENZIRREV
Ala_ex Arg_ex Asn_ex Asp_ex Cys_ex G3P_up GC_1 GC_2 GNC Gln_ex Glu_ex Gly_ex ORC_1 PYC
Pro_ex Pyr_up R00149 R00209 R00239 R00253 R00256 R00342 R00351 R00369 R00431 R00551 R00557
R00578 R00582 R00621 R00782 R00893 R00896 R01086 R01398 R01954 R02619 R03105 R03316 R03425
R03815 R04125 R04861 Ser_ex SufS

-METINT
(S)-1-Pyrroline-5-carboxylate_mit (S)-Malate_cyt (S)-Malate_mit 2-Oxoglutarate_cyt
2-Oxoglutarate_mit 3-Carboxy-1-hydroxypropyl-ThPP_mit 3-Phospho-D-glycerate_cyt
3-Phosphonooxypyruvate_cyt 3-Sulfino-L-alanine_cyt 3-Sulfinylpyruvate_cyt Acetyl-CoA_mit
C01242_mit Carbamoyl-P_mit Citrate_mit Dihydrolipoamide_mit Dihydrolipoylprotein_mit
Fumarate_cyt Fumarate_mit Glyceraldehyde-3P_cyt Glycerate-1,3P2_cyt Glycine_cyt
Glycine_mit Isocitrate_mit L-Alanine_cyt L-Alanine_mit L-Arginine_cyt L-Asparagine_cyt
L-Aspartate_cyt L-Citrulline_cyt L-Citrulline_mit L-Cysteine_cyt L-Cysteine_mit
L-Glutamate_5-semialdehyde_mit L-Glutamate_cyt L-Glutamate_mit L-Glutamine_cyt
L-Glutamine_mit L-Glutamyl-P_mit L-Ornithine_cyt L-Ornithine_mit L-Proline_cyt
L-Proline_mit L-Serine_cyt L-Serine_mit Lipoamide_mit Lipoylprotein_mit
Mercaptopyruvate_cyt N-(L-Arginino)succinate_cyt O-Phospho-L-serine_cyt Oxaloacetate_cyt
Oxaloacetate_mit Oxalosuccinate_mit Phosphoenolpyruvate_cyt Pyruvate_cyt Pyruvate_mit
S-Succinyldihydrolipoamide_mit Succinate_mit Succinyl-CoA_mit ThPP_mit

-METEXT
5,10-Methylene-THF_mit ADP_cyt ADP_mit AMP_cyt ATP_cyt ATP_mit CO2_cyt CO2_mit CoA_mit
GDP_cyt GDP_mit GTP_cyt GTP_mit Glyceraldehyde-3P_ext Glycine_ext Glyoxylate_cyt H+_cyt
H+_mit H2O_cyt H2O_mit H2S2O3_cyt H2S_cyt L-Alanine_ext L-Arginine_ext L-Asparagine_ext
L-Aspartate_ext L-Cysteine_ext L-Glutamate_ext L-Glutamine_ext L-Proline_ext L-Serine_ext
NAD+_cyt NAD+_mit NADH_cyt NADH_mit NADP+_cyt NADP+_mit NADPH_cyt NADPH_mit NH3_cyt
NH3_mit NO_cyt O2_cyt OH-_cyt OH-_mit PPi_cyt Pi_cyt Pi_mit Pyruvate_ext SO2_cyt
Succinate_cyt Sulfite_cyt THF_mit Ubiquinol_mit Ubiquinone_mit Urea_cyt
[enzyme]-S-sulfanylcysteine_mit [enzyme]-cysteine_mit

-CAT
Ala_diff : L-Alanine_cyt = L-Alanine_mit .
Ala_ex : L-Alanine_cyt = L-Alanine_ext .
Arg_ex : L-Arginine_cyt = L-Arginine_ext .
Asn_ex : L-Asparagine_cyt = L-Asparagine_ext .
Asp_ex : L-Aspartate_cyt = L-Aspartate_ext .
Cys_diff : L-Cysteine_cyt = L-Cysteine_mit .
Cys_ex : L-Cysteine_cyt = L-Cysteine_ext .
DIC_1 : Pi_cyt + (S)-Malate_mit = Pi_mit + (S)-Malate_cyt .
DIC_2 : Succinate_cyt + Fumarate_mit = Succinate_mit + Fumarate_cyt .
G3P_up : Glyceraldehyde-3P_ext = Glyceraldehyde-3P_cyt .
GC_1 : L-Glutamate_cyt + H+_cyt = L-Glutamate_mit + H+_mit .
GC_2 : L-Glutamate_cyt + OH-_mit = L-Glutamate_mit + OH-_cyt .
```

```
GNC : L-Glutamine_cyt + H+_cyt = L-Glutamine_mit + H+_mit .
Gln_ex : L-Glutamine_cyt = L-Glutamine_ext .
Glu_ex : L-Glutamate_cyt = L-Glutamate_ext .
Gly_diff : Glycine_cyt = Glycine_mit .
Gly_ex : Glycine_cyt = Glycine_ext .
OGC : (S)-Malate_cyt + 2-Oxoglutarate_mit = (S)-Malate_mit + 2-Oxoglutarate_cyt .
ORC_1 : L-Ornithine_cyt + L-Citrulline_mit + H+_mit = L-Ornithine_mit +
        L-Citrulline_cyt + H+_cyt .
ORC_2 : L-Ornithine_cyt + H+_mit = L-Ornithine_mit + H+_cyt .
PYC : Pyruvate_cyt + H+_cyt = Pyruvate_mit + H+_mit .
Pro_diff : L-Proline_cyt = L-Proline_mit .
Pro_ex : L-Proline_cyt = L-Proline_ext .
Pro_tr : L-Proline_cyt + L-Glutamate_mit = L-Proline_mit + L-Glutamate_cyt .
Pyr_up : Pyruvate_ext = Pyruvate_cyt .
R00149 : 2 ATP_mit + NH3_mit + CO2_mit + H2O_mit = 2 ADP_mit + Pi_mit + Carbamoyl-P_mit .
R00209 : Pyruvate_mit + CoA_mit + NAD+_mit = Acetyl-CoA_mit + CO2_mit + NADH_mit .
R00214 : (S)-Malate_mit + NAD+_mit = Pyruvate_mit + CO2_mit + NADH_mit .
R00239 : ATP_mit + L-Glutamate_mit = ADP_mit + L-Glutamyl-P_mit .
R00243 : L-Glutamate_mit + NAD+_mit + H2O_mit = 2-Oxoglutarate_mit + NH3_mit +
         NADH_mit + H+_mit .
R00245 : L-Glutamate_5-semialdehyde_mit + NAD+_mit + H2O_mit = L-Glutamate_mit +
         NADH_mit + H+_mit .
R00253 : ATP_cyt + L-Glutamate_cyt + NH3_cyt = ADP_cyt + Pi_cyt + L-Glutamine_cyt .
R00256 : L-Glutamine_mit + H2O_mit = L-Glutamate_mit + NH3_mit .
R00258 : L-Alanine_mit + 2-Oxoglutarate_mit = Pyruvate_mit + L-Glutamate_mit .
R00268 : Oxalosuccinate_mit = 2-Oxoglutarate_mit + CO2_mit .
R00342 : (S)-Malate_cyt + NAD+_cyt = Oxaloacetate_cyt + NADH_cyt + H+_cyt .
R00344 : ATP_mit + Pyruvate_mit + CO2_mit + H2O_mit = ADP_mit + Pi_mit +
         Oxaloacetate_mit .
R00351 : Acetyl-CoA_mit + H2O_mit + Oxaloacetate_mit = Citrate_mit + CoA_mit .
R00355 : L-Aspartate_cyt + 2-Oxoglutarate_cyt = Oxaloacetate_cyt + L-Glutamate_cyt .
R00369 : L-Alanine_cyt + Glyoxylate_cyt = Pyruvate_cyt + Glycine_cyt .
R00405 : ATP_mit + Succinate_mit + CoA_mit = ADP_mit + Pi_mit + Succinyl-CoA_mit .
R00430 : GTP_cyt + Pyruvate_cyt = GDP_cyt + Phosphoenolpyruvate_cyt .
R00431 : GTP_cyt + Oxaloacetate_cyt = GDP_cyt + Phosphoenolpyruvate_cyt + CO2_cyt .
R00432 : GTP_mit + Succinate_mit + CoA_mit = GDP_mit + Pi_mit + Succinyl-CoA_mit .
R00551 : L-Arginine_cyt + H2O_cyt = L-Ornithine_cyt + Urea_cyt .
R00557 : L-Arginine_cyt + O2_cyt + NADPH_cyt + H+_cyt = NO_cyt + L-Citrulline_cyt +
         NADP+_cyt .
R00578 : ATP_cyt + L-Aspartate_cyt + L-Glutamine_cyt + H2O_cyt = AMP_cyt + PPi_cyt +
         L-Asparagine_cyt + L-Glutamate_cyt .
R00582 : O-Phospho-L-serine_cyt + H2O_cyt = L-Serine_cyt + Pi_cyt .
R00621 : 2-Oxoglutarate_mit + ThPP_mit = 3-Carboxy-1-hydroxypropyl-ThPP_mit + CO2_mit .
R00667 : L-Ornithine_mit + 2-Oxoglutarate_mit = L-Glutamate_5-semialdehyde_mit +
         L-Glutamate_mit .
R00707 : (S)-1-Pyrroline-5-carboxylate_mit + NAD+_mit + 2 H2O_mit = L-Glutamate_mit +
         NADH_mit + H+_mit .
R00709 : Isocitrate_mit + NAD+_mit = 2-Oxoglutarate_mit + CO2_mit + NADH_mit + H+_mit .
R00782 : H2S_cyt + Pyruvate_cyt + NH3_cyt = L-Cysteine_cyt + H2O_cyt .
R00893 : L-Cysteine_cyt + O2_cyt = 3-Sulfino-L-alanine_cyt .
R00896 : Mercaptopyruvate_cyt + L-Glutamate_cyt = L-Cysteine_cyt + 2-Oxoglutarate_cyt .
```

```
R00945 : 5,10-Methylene-THF_mit + Glycine_mit + H2O_mit = THF_mit + L-Serine_mit .
R01061 : Glyceraldehyde-3P_cyt + Pi_cyt + NAD+_cyt = Glycerate-1,3P2_cyt +
         NADH_cyt + H+_cyt .
R01082 : (S)-Malate_mit = Fumarate_mit + H2O_mit .
R01086 : N-(L-Arginino)succinate_cyt = Fumarate_cyt + L-Arginine_cyt .
R01248 : L-Proline_mit + NAD+_mit = (S)-1-Pyrroline-5-carboxylate_mit +
         NADH_mit + H+_mit .
R01324 : Citrate_mit = Isocitrate_mit .
R01398 : Carbamoyl-P_mit + L-Ornithine_mit = Pi_mit + L-Citrulline_mit .
R01512 : ATP_cyt + 3-Phospho-D-glycerate_cyt = ADP_cyt + Glycerate-1,3P2_cyt .
R01513 : 3-Phospho-D-glycerate_cyt + NAD+_cyt = 3-Phosphonooxypyruvate_cyt +
         NADH_cyt + H+_cyt .
R01698 : Dihydrolipoamide_mit + NAD+_mit = Lipoamide_mit + NADH_mit + H+_mit .
R01899 : Isocitrate_mit + NADP+_mit = Oxalosuccinate_mit + NADPH_mit + H+_mit .
R01954 : ATP_cyt + L-Citrulline_cyt + L-Aspartate_cyt = AMP_cyt + PPi_cyt +
         N-(L-Arginino)succinate_cyt .
R02164 : Ubiquinone_mit + Succinate_mit = Ubiquinol_mit + Fumarate_mit .
R02570 : Succinyl-CoA_mit + Dihydrolipoamide_mit = CoA_mit +
         S-Succinyldihydrolipoamide_mit .
R02619 : 3-Sulfino-L-alanine_cyt + 2-Oxoglutarate_cyt = 3-Sulfinylpyruvate_cyt +
         L-Glutamate_cyt .
R03105 : H2S2O3_cyt + Pyruvate_cyt = Mercaptopyruvate_cyt + Sulfite_cyt .
R03313 : L-Glutamate_5-semialdehyde_mit + Pi_mit + NADP+_mit = L-Glutamyl-P_mit +
         NADPH_mit + H+_mit .
R03316 : 3-Carboxy-1-hydroxypropyl-ThPP_mit + Lipoamide_mit =
         S-Succinyldihydrolipoamide_mit + ThPP_mit .
R03425 : Glycine_mit + Lipoylprotein_mit = C01242_mit + CO2_mit .
R03815 : Dihydrolipoylprotein_mit + NAD+_mit = Lipoylprotein_mit + NADH_mit + H+_mit .
R04125 : THF_mit + C01242_mit = 5,10-Methylene-THF_mit + NH3_mit +
         Dihydrolipoylprotein_mit .
R04173 : O-Phospho-L-serine_cyt + 2-Oxoglutarate_cyt = 3-Phosphonooxypyruvate_cyt +
         L-Glutamate_cyt .
R04861 : 3-Sulfinylpyruvate_cyt = SO2_cyt + Pyruvate_cyt .
Ser_diff : L-Serine_cyt = L-Serine_mit .
Ser_ex : L-Serine_cyt = L-Serine_ext .
SufS : L-Cysteine_mit + [enzyme]-cysteine_mit = L-Alanine_mit +
       [enzyme]-S-sulfanylcysteine_mit .
spont2 : L-Glutamate_5-semialdehyde_mit = (S)-1-Pyrroline-5-carboxylate_mit + H2O_mit .
```

A.5 Supplement 5

Additional transport reactions found during literature search for the `Metatool` model of the human hepatocyte amino acid system (see Supplement 4). These transport reactions were not used in our hepatocyte amino acid anabolism but contributed to the comprehensive hepatocyte metabolism *HepatoNet1* that was reconstructed in the group of Prof. Holzhütter at the Charité in Berlin (see Gille *et al.*, 2010).
Identifiers have the same meaning as given in Supplement 1 but METINT and METEXT are missing since the following list of reactions is not a complete `Metatool` input file.

```
-ENZREV
AGC Ac_diff Arg_tr CIC_1 CIC_2 CO2_diff Citrul_tr For_diff GDC H2O_diff HCO3_diff Ile_tr
Leu_tr Met_tr NH3_diff NO3_diff ODC Tcynt_diff Thr_tr Val_tr

-ENZIRREV
AAC APC Arg_tr2 CAC PiC_1

-CAT
AAC : ADP_cyt + ATP_mit = ADP_mit + ATP_cyt .
AGC : L-Glutamate_cyt + H+_cyt + L-Aspartate_mit = L-Glutamate_mit + H+_mit +
      L-Aspartate_cyt .
APC : ATP_cyt + Pi_mit = ATP_mit + Pi_cyt .
Ac_diff : Acetate_cyt = Acetate_mit .
Arg_tr : L-Arginine_cyt  = L-Arginine_mit .
Arg_tr2 : L-Arginine_cyt + H+_cyt = L-Arginine_mit + H+_mit .
CAC : O-Acylcarnitine_cyt + L-Carnitine_mit = O-Acylcarnitine_mit + L-Carnitine_cyt .
CIC_1 : (S)-Malate_cyt + Citrate_mit = (S)-Malate_mit + Citrate_cyt .
CIC_2 : (S)-Malate_cyt + Isocitrate_mit = (S)-Malate_mit + Isocitrate_cyt .
CO2_diff : CO2_cyt = CO2_mit .
Citrul_tr : L-Citrulline_cyt = L-Citrulline_mit .
For_diff : Formate_cyt = Formate_mit .
GDC : CoA_cyt = CoA_mit .
H2O_diff : H2O_cyt = H2O_mit .
HCO3_diff : Hydrogencarbonate_cyt = Hydrogencarbonate_mit .
Ile_tr : L-Isoleucine_cyt = L-Isoleucine_mit .
Leu_tr : L-Leucine_cyt = L-Leucine_mit .
Met_tr : L-Methionine_cyt = L-Methionine_mit .
NH3_diff : NH3_cyt = NH3_mit .
NO3_diff : Nitrate_cyt = Nitrate_mit .
ODC : 2-Oxoadipate_cyt + 2-Oxoglutarate_mit = 2-Oxoadipate_mit + 2-Oxoglutarate_cyt .
PiC_1 : Pi_cyt + H+_cyt = Pi_mit + H+_mit .
Tcynt_diff : Thiocyanate_cyt = Thiocyanate_mit .
Thr_tr : L-Threonine_cyt = L-Threonine_mit .
Val_tr : L-Valine_cyt = L-Valine_mit .
```

A.6 Supplement 6

Robustness values plotted in Figure 1 in the main text

knockout depth d resp. D	$R_1(d)$	$R_1(\leq D)$
1	0.4631	0.4631
2	0.2230	0.3997
3	0.1116	0.3860
4	0.0577	0.3837
5	0.0307	0.3833
6	0.0167	0.3833
7	0.0092	0.3833
8	0.0051	0.3833
9	0.0028	0.3833
10	0.0015	0.3833

B Supplementary Material to Chapter 4

B.1 Supplement 1a

Input file for the program `Metatool` of the single phosphorylation cascade, comprising the kinases E_1–E_5.
Identifiers: ENZREV, reversible enzymes; ENZIRREV, irreversible enzymes; METINT, internal metabolites; METEXT, external metabolites; CAT, catalysed reactions.

```
-ENZREV


-ENZIRREV
R1 R2 R3 R4

-METINT
E2 E2P E3 E3P E4 E4P

-METEXT
E1 E1P E5 E5P NDP NTP P

-CAT
R1 : 1 NTP + 1 E2 + 1 E1P = 1 E2P + 1 NDP + 1 E1 + 1 P
R2 : 1 NTP + 1 E3 + 1 E2P = 1 E3P + 1 NDP + 1 E2 + 1 P
R3 : 1 NTP + 1 E4 + 1 E3P = 1 E4P + 1 NDP + 1 E3 + 1 P
R4 : 1 NTP + 1 E5 + 1 E4P = 1 E5P + 1 NDP + 1 E4 + 1 P
```

B.2 Supplement 1b

Output file for the program `Metatool` of the single phosphorylation cascade, comprising the kinases E_1–E_5.

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 4.804 * edges^(-0.2075)
Linear correlation coefficient: -0.4131.
The dependency is not significant (p > 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r6 x c4
```

```
-1 1 0 0
1 -1 0 0
0 -1 1 0
0 1 -1 0
0 0 -1 1
0 0 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1


KERNEL

 matrix dimension r1 x c4
-1 -1 -1 -1

 enzymes

 1: (4) -R1 -R2 -R3 -R4

 overall reaction

 1: E1 + E5P + 4 NDP + 4 P = E1P + E5 + 4 NTP


CONSERVATION RELATIONS

 matrix dimension r3 x c6
1 1 -0 0 -0 0
-0 0 1 1 -0 0
-0 0 -0 0 1 1

 1: E2 + E2P = const
 2: E3 + E3P = const
 3: E4 + E4P = const


SUBSETS

 matrix dimension r1 x c4
1 1 1 1

 enzymes

 1: (4) R1 R2 R3 R4 irreversible

 overall reaction

 1: E1P + E5 + 4 NTP = E1 + E5P + 4 NDP + 4 P


REDUCED SYSTEM
```

```
 matrix dimension r1 x c1
0
The following line indicates reversible (0) and irreversible reactions (1)
1


ELEMENTARY MODES

 matrix dimension r1 x c4
1 1 1 1

 enzymes

 1: (4) R1 R2 R3 R4 irreversible

 overall reaction

 1: E1P + E5 + 4 NTP = E1 + E5P + 4 NDP + 4 P
```

B.3 Supplement 2a

Input file for the program `Metatool` of the OR-connected diverging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3 is given in Table 1 in the main article.

B.4 Supplement 2b

Output file for the program `Metatool` of the OR-connected diverging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 6.836 * edges^(-0.4656)
Linear correlation coefficient: -0.565.
The dependency is not significant (p > 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r12 x c8
-1 1 0 0 0 0 0 0
1 -1 0 0 0 0 0 0
0 -1 1 0 0 1 0 0
0 1 -1 0 0 -1 0 0
```

```
0 0 -1 1 0 0 0 0
0 0 1 -1 0 0 0 0
0 0 0 -1 1 0 0 0
0 0 0 1 -1 0 0 0
0 0 0 0 0 -1 1 0
0 0 0 0 0 1 -1 0
0 0 0 0 0 0 -1 1
0 0 0 0 0 0 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1 1 1 1 1


KERNEL

 matrix dimension r2 x c8
-1 -1 -1 -1 -1 -0 -0 -0
-1 -1 -0 -0 -0 -1 -1 -1

 enzymes

 1: (5) -R1 -R2 -R3 -R4 -R5
 2: (5) -R1 -R2 -R6 -R7 -R8

 overall reaction

 1: E1 + F3P + 5 NDP + 5 P = E1P + F3 + 5 NTP
 2: E1 + G3P + 5 NDP + 5 P = E1P + G3 + 5 NTP


CONSERVATION RELATIONS

 matrix dimension r6 x c12
1 1 -0 0 -0 0 -0 0 -0 0 -0 0
-0 0 1 1 -0 0 -0 0 -0 0 -0 0
-0 0 -0 0 1 1 -0 0 -0 0 -0 0
-0 0 -0 0 -0 0 1 1 -0 0 -0 0
-0 0 -0 0 -0 0 -0 0 1 1 -0 0
-0 0 -0 0 -0 0 -0 0 -0 0 1 1

 1: E2 + E2P = const
 2: E3 + E3P = const
 3: F1 + F1P = const
 4: F2 + F2P = const
 5: G1 + G1P = const
 6: G2 + G2P = const


SUBSETS

 matrix dimension r3 x c8
1 1 0 0 0 0 0 0
```

```
0 0 1 1 1 0 0 0
0 0 0 0 0 1 1 1

 enzymes

 1: (2) R1 R2 irreversible
 2: (3) R3 R4 R5 irreversible
 3: (3) R6 R7 R8 irreversible

 overall reaction

 1: E3 + E1P + 2 NTP = E3P + E1 + 2 NDP + 2 P
 2: E3P + F3 + 3 NTP = E3 + F3P + 3 NDP + 3 P
 3: E3P + G3 + 3 NTP = E3 + G3P + 3 NDP + 3 P


REDUCED SYSTEM

 matrix dimension r2 x c3
-1 1 1
1 -1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1


ELEMENTARY MODES

 matrix dimension r2 x c8
1 1 1 1 1 0 0 0
1 1 0 0 0 1 1 1

 enzymes

 1: (5) R1 R2 R3 R4 R5 irreversible
 2: (5) R1 R2 R6 R7 R8 irreversible

 overall reaction

 1: E1P + F3 + 5 NTP = E1 + F3P + 5 NDP + 5 P
 2: E1P + G3 + 5 NTP = E1 + G3P + 5 NDP + 5 P
```

B.5 Supplement 3a

Input file for the program `Metatool` of the AND-connected diverging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3 Identifiers are as in Supplement 1a.

```
-ENZREV


-ENZIRREV
R1 R2 R3 R4 R5 R6 R7

-METINT
E2 E2P E3 E3P F1 F1P F2 F2P G1 G1P G2 G2P

-METEXT
E1 E1P F3 F3P G3 G3P NDP NTP P

-CAT
R1 : 1 NTP + 1 E2 + 1 E1P = 1 E2P + 1 NDP + 1 E1 + 1 P
R2 : 1 NTP + 1 E3 + 1 E2P = 1 E3P + 1 NDP + 1 E2 + 1 P
R3 : 2 NTP + 1 F1 + 1 G1 + 1 E3P = 1 F1P + 1 G1P + 2 NDP + 1 E3 + 1 P
R4 : 1 NTP + 1 F2 + 1 F1P = 1 F2P + 1 NDP + 1 F1 + 1 P
R5 : 1 NTP + 1 F3 + 1 F2P = 1 F3P + 1 NDP + 1 F2 + 1 P
R6 : 1 NTP + 1 G2 + 1 G1P = 1 G2P + 1 NDP + 1 G1 + 1 P
R7 : 1 NTP + 1 G3 + 1 G2P = 1 G3P + 1 NDP + 1 G2 + 1 P
```

B.6 Supplement 3b

Output file for the program `Metatool` of the AND-connected diverging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 8.885 * edges^(-0.4463)
Linear correlation coefficient: -0.6351.
The dependency is not significant (p > 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r12 x c7
-1 1 0 0 0 0 0
1 -1 0 0 0 0 0
0 -1 1 0 0 0 0
0 1 -1 0 0 0 0
0 0 -1 1 0 0 0
0 0 1 -1 0 0 0
0 0 0 -1 1 0 0
0 0 0 1 -1 0 0
0 0 -1 0 0 1 0
0 0 1 0 0 -1 0
0 0 0 0 0 -1 1
```

```
0 0 0 0 0 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1 1 1 1


KERNEL

 matrix dimension r1 x c7
-1 -1 -1 -1 -1 -1 -1

 enzymes

 1: (7) -R1 -R2 -R3 -R4 -R5 -R6 -R7

 overall reaction

 1: E1 + F3P + G3P + 8 NDP + 7 P = E1P + F3 + G3 + 8 NTP


CONSERVATION RELATIONS

 matrix dimension r6 x c12
1 1 -0 0 -0 0 -0 0 -0 0 -0 0
-0 0 1 1 -0 0 -0 0 -0 0 -0 0
-0 0 -0 0 1 1 -0 0 -0 0 -0 0
-0 0 -0 0 -0 0 1 1 -0 0 -0 0
-0 0 -0 0 -0 0 -0 0 1 1 -0 0
-0 0 -0 0 -0 0 -0 0 -0 0 1 1

 1: E2 + E2P = const
 2: E3 + E3P = const
 3: F1 + F1P = const
 4: F2 + F2P = const
 5: G1 + G1P = const
 6: G2 + G2P = const


SUBSETS

 matrix dimension r1 x c7
1 1 1 1 1 1 1

 enzymes

 1: (7) R1 R2 R3 R4 R5 R6 R7 irreversible

 overall reaction

 1: E1P + F3 + G3 + 8 NTP = E1 + F3P + G3P + 8 NDP + 7 P
```

```
REDUCED SYSTEM

 matrix dimension r1 x c1
0
The following line indicates reversible (0) and irreversible reactions (1)
1


ELEMENTARY MODES

 matrix dimension r1 x c7
1 1 1 1 1 1 1

 enzymes

 1: (7) R1 R2 R3 R4 R5 R6 R7 irreversible

 overall reaction

 1: E1P + F3 + G3 + 8 NTP = E1 + F3P + G3P + 8 NDP + 7 P
```

B.7 Supplement 4a

Input file for the program `Metatool` of the OR-connected converging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3 is given in Table 2 in the main article.

B.8 Supplement 4b

Output file for the program `Metatool` of the OR-connected converging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 6.836 * edges^(-0.4656)
Linear correlation coefficient: -0.565.
The dependency is not significant (p > 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r12 x c8
-1 1 0 0 0 0 0 0
1 -1 0 0 0 0 0 0
0 -1 0 0 1 0 0 0
```

```
0 1 0 0 -1 0 0 0
0 0 -1 1 0 0 0 0
0 0 1 -1 0 0 0 0
0 0 0 -1 0 1 0 0
0 0 0 1 0 -1 0 0
0 0 0 0 -1 -1 1 0
0 0 0 0 1 1 -1 0
0 0 0 0 0 0 -1 1
0 0 0 0 0 0 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1 1 1 1 1


KERNEL

 matrix dimension r2 x c8
1 1 -1 -1 1 -1 0 0
-1 -1 -0 -0 -1 -0 -1 -1

 enzymes

 1: (6) R1 R2 -R3 -R4 R5 -R6
 2: (5) -R1 -R2 -R5 -R7 -R8

 overall reaction

 1: E1P + F1 = E1 + F1P
 2: E1 + G3P + 5 NDP + 5 P = E1P + G3 + 5 NTP


CONSERVATION RELATIONS

 matrix dimension r6 x c12
1 1 -0 0 -0 0 -0 0 -0 0 -0 0
-0 0 1 1 -0 0 -0 0 -0 0 -0 0
-0 0 -0 0 1 1 -0 0 -0 0 -0 0
-0 0 -0 0 -0 0 1 1 -0 0 -0 0
-0 0 -0 0 -0 0 -0 0 1 1 -0 0
-0 0 -0 0 -0 0 -0 0 -0 0 1 1

 1: E2 + E2P = const
 2: E3 + E3P = const
 3: F2 + F2P = const
 4: F3 + F3P = const
 5: G1 + G1P = const
 6: G2 + G2P = const


SUBSETS

 matrix dimension r3 x c8
```

```
1 1 0 0 1 0 0 0
0 0 1 1 0 1 0 0
0 0 0 0 0 0 1 1

 enzymes

 1: (3) R1 R2 R5 irreversible
 2: (3) R3 R4 R6 irreversible
 3: (2) R7 R8 irreversible

 overall reaction

 1: G1 + E1P + 3 NTP = G1P + E1 + 3 NDP + 3 P
 2: G1 + F1P + 3 NTP = G1P + F1 + 3 NDP + 3 P
 3: G1P + G3 + 2 NTP = G1 + G3P + 2 NDP + 2 P


REDUCED SYSTEM

 matrix dimension r2 x c3
-1 -1 1
1 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1


ELEMENTARY MODES

 matrix dimension r2 x c8
0 0 1 1 0 1 1 1
1 1 0 0 1 0 1 1

 enzymes

 1: (5) R3 R4 R6 R7 R8 irreversible
 2: (5) R1 R2 R5 R7 R8 irreversible

 overall reaction

 1: F1P + G3 + 5 NTP = F1 + G3P + 5 NDP + 5 P
 2: E1P + G3 + 5 NTP = E1 + G3P + 5 NDP + 5 P
```

B.9 Supplement 5a

Input file for the program `Metatool` of the AND-connected converging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3 Identifiers are as in Supplement 1a.

```
-ENZREV


-ENZIRREV
R1 R2 R3 R4 R5 R6 R7

-METINT
E2 E2P E3 E3P F2 F2P F3 F3P G1 G1P G2 G2P

-METEXT
E1 E1P F1 F1P G3 G3P NDP NTP P

-CAT
R1 : 1 NTP + 1 E2 + 1 E1P = 1 E2P + 1 NDP + 1 E1 + 1 P
R2 : 1 NTP + 1 E3 + 1 E2P = 1 E3P + 1 NDP + 1 E2 + 1 P
R3 : 1 NTP + 1 F2 + 1 F1P = 1 F2P + 1 NDP + 1 F1 + 1 P
R4 : 1 NTP + 1 F3 + 1 F2P = 1 F3P + 1 NDP + 1 F2 + 1 P
R5 : 1 NTP + 1 G1 + 1 E3P + 1 F3P = 1 G1P + 1 NDP + 1 E3 + 1 F3 + 2 P
R6 : 1 NTP + 1 G2 + 1 G1P = 1 G2P + 1 NDP + 1 G1 + 1 P
R7 : 1 NTP + 1 G3 + 1 G2P = 1 G3P + 1 NDP + 1 G2 + 1 P
```

B.10 Supplement 5b

Output file for the program `Metatool` of the AND-connected converging single phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 8.885 * edges^(-0.4463)
Linear correlation coefficient: -0.6351.
The dependency is not significant (p > 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r12 x c7
-1 1 0 0 0 0 0
1 -1 0 0 0 0 0
0 -1 0 0 1 0 0
0 1 0 0 -1 0 0
0 0 -1 1 0 0 0
0 0 1 -1 0 0 0
0 0 0 -1 1 0 0
0 0 0 1 -1 0 0
0 0 0 0 -1 1 0
```

```
0 0 0 0 1 -1 0
0 0 0 0 0 -1 1
0 0 0 0 0 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1 1 1 1


KERNEL

 matrix dimension r1 x c7
-1 -1 -1 -1 -1 -1 -1

 enzymes

 1: (7) -R1 -R2 -R3 -R4 -R5 -R6 -R7

 overall reaction

 1: E1 + F1 + G3P + 7 NDP + 8 P = E1P + F1P + G3 + 7 NTP


CONSERVATION RELATIONS

 matrix dimension r6 x c12
1 1 -0 0 -0 0 -0 0 -0 0 -0 0
-0 0 1 1 -0 0 -0 0 -0 0 -0 0
-0 0 -0 0 1 1 -0 0 -0 0 -0 0
-0 0 -0 0 -0 0 1 1 -0 0 -0 0
-0 0 -0 0 -0 0 -0 0 1 1 -0 0
-0 0 -0 0 -0 0 -0 0 -0 0 1 1

 1: E2 + E2P = const
 2: E3 + E3P = const
 3: F2 + F2P = const
 4: F3 + F3P = const
 5: G1 + G1P = const
 6: G2 + G2P = const


SUBSETS

 matrix dimension r1 x c7
1 1 1 1 1 1 1

 enzymes

 1: (7) R1 R2 R3 R4 R5 R6 R7 irreversible

 overall reaction

 1: E1P + F1P + G3 + 7 NTP = E1 + F1 + G3P + 7 NDP + 8 P
```

```
REDUCED SYSTEM

 matrix dimension r1 x c1
0
The following line indicates reversible (0) and irreversible reactions (1)
1


ELEMENTARY MODES

 matrix dimension r1 x c7
1 1 1 1 1 1 1

 enzymes

 1: (7) R1 R2 R3 R4 R5 R6 R7 irreversible

 overall reaction

 1: E1P + F1P + G3 + 7 NTP = E1 + F1 + G3P + 7 NDP + 8 P
```

B.11 Supplement 6a

Input file for the program `Metatool` of the simplified double phosphorylation cascade, comprising the kinases E_1–E_4 is given in Table 3 in the main article.

B.12 Supplement 6b

Output file for the program `Metatool` of the simplified double phosphorylation cascade, comprising the kinases E_1–E_4

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 4.179 * edges^(-0.2337)
Linear correlation coefficient: -0.7816.
The dependency is significant (p < 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r4 x c3
-1 1 0
1 -1 0
```

```
0 -1 1
0 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1


KERNEL

 matrix dimension r1 x c3
-1 -1 -1

 enzymes

 1: (3) -R1 -R2 -R3

 overall reaction

 1: E1 + E4PP + 6 NDP + 6 P = E1PP + E4 + 6 NTP


CONSERVATION RELATIONS

 matrix dimension r2 x c4
1 1 -0 0
-0 0 1 1

 1: E2 + E2PP = const
 2: E3 + E3PP = const


SUBSETS

 matrix dimension r1 x c3
1 1 1

 enzymes

 1: (3) R1 R2 R3 irreversible

 overall reaction

 1: E1PP + E4 + 6 NTP = E1 + E4PP + 6 NDP + 6 P


REDUCED SYSTEM

 matrix dimension r1 x c1
0
The following line indicates reversible (0) and irreversible reactions (1)
1
```

```
ELEMENTARY MODES

 matrix dimension r1 x c3
1 1 1

 enzymes

 1: (3) R1 R2 R3 irreversible

 overall reaction

 1: E1PP + E4 + 6 NTP = E1 + E4PP + 6 NDP + 6 P
```

B.13 Supplement 7a

Input file for the program `Metatool` of the simplified diverging double-phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3 is given in Table 4 in the main article.

B.14 Supplement 7b

Output file for the program `Metatool` of the simplified diverging double-phosphorylation cascade, comprising the kinases E_1–E_3, F_1–F_3 and G_1–G_3

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 6.09 * edges^(-0.5855)
Linear correlation coefficient: -0.5279.
The dependency is not significant (p > 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r13 x c9
-1 1 1 0 0 0 0 0 0
1 -1 -1 0 0 0 0 0 0
0 -1 -1 1 1 0 0 0 0
0 1 0 -1 0 0 0 0 0
0 0 1 0 -1 0 0 0 0
0 0 0 -1 0 1 0 0 0
0 0 0 1 0 -1 0 0 0
0 0 0 0 0 -1 1 0 0
0 0 0 0 0 1 -1 0 0
0 0 0 0 -1 0 0 1 0
```

```
0 0 0 0 1 0 0 -1 0
0 0 0 0 0 0 0 -1 1
0 0 0 0 0 0 0 1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1 1 1 1 1 1


KERNEL

 matrix dimension r2 x c9
-1 -1 0 -1 0 -1 -1 0 0
-1 0 -1 0 -1 0 0 -1 -1

 enzymes

 1: (5) -R1 -R2 -R4 -R6 -R7
 2: (5) -R1 -R3 -R5 -R8 -R9

 overall reaction

 1: E1 + F3PP + 9 NDP + 9 P = E1PP + F3 + 9 NTP
 2: E1 + G3PP + 10 NDP + 10 P = E1PP + G3 + 10 NTP


CONSERVATION RELATIONS

 matrix dimension r6 x c13
1 1 -0 -0 0 -0 0 -0 0 -0 0 -0 0
-0 0 1 1 1 -0 0 -0 0 -0 0 -0 0
-0 0 -0 -0 0 1 1 -0 0 -0 0 -0 0
-0 0 -0 -0 0 -0 0 1 1 -0 0 -0 0
-0 0 -0 -0 0 -0 0 -0 0 1 1 -0 0
-0 0 -0 -0 0 -0 0 -0 0 -0 0 1 1

 1: E2 + E2PP = const
 2: E3 + E3P + E3PP = const
 3: F1 + F1PP = const
 4: F2 + F2PP = const
 5: G1 + G1PP = const
 6: G2 + G2PP = const


SUBSETS

 matrix dimension r3 x c9
1 0 0 0 0 0 0 0 0
0 1 0 1 0 1 1 0 0
0 0 1 0 1 0 0 1 1

 enzymes
```

```
 1: (1) R1 irreversible
 2: (4) R2 R4 R6 R7 irreversible
 3: (4) R3 R5 R8 R9 irreversible

 overall reaction

 1: E2 + E1PP + 2 NTP = E2PP + E1 + 2 NDP + 2 P
 2: E2PP + F3 + 7 NTP = E2 + F3PP + 7 NDP + 7 P
 3: E2PP + G3 + 8 NTP = E2 + G3PP + 8 NDP + 8 P


REDUCED SYSTEM

 matrix dimension r2 x c3
-1 1 1
1 -1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1


ELEMENTARY MODES

 matrix dimension r2 x c9
1 1 0 1 0 1 1 0 0
1 0 1 0 1 0 0 1 1

 enzymes

 1: (5) R1 R2 R4 R6 R7 irreversible
 2: (5) R1 R3 R5 R8 R9 irreversible

 overall reaction

 1: E1PP + F3 + 9 NTP = E1 + F3PP + 9 NDP + 9 P
 2: E1PP + G3 + 10 NTP = E1 + G3PP + 10 NDP + 10 P
```

B.15 Supplement 8a

Input file for the program `Metatool` of the modelled part of the insulin pathway is given in Table 5 in the main article.

B.16 Supplement 8b

Output file for the program `Metatool` of the modelled part of the insulin pathway

```
METATOOL OUTPUT Version 5.1

freq_of_nodes = 16.43 * edges^(-1.214)
Linear correlation coefficient: -0.8096.
The dependency is significant (p < 0.05).


STOICHIOMETRIC MATRIX

 matrix dimension r22 x c15
0 0 0 0 0 -1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 1 -1 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 -1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 -1 0
-1 0 0 0 0 0 0 0 0 0 1 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 -1 0 0 0
0 -1 0 0 0 0 0 0 0 0 0 1 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 -1 0
0 0 -1 0 0 0 0 0 0 0 0 0 0 1 0
0 0 1 0 0 0 0 0 0 0 -1 0 0 0 0
0 0 1 0 0 0 0 0 0 0 -1 0 0 0 0
0 0 0 0 -1 1 0 0 0 0 0 0 0 0 0
0 0 0 0 1 -1 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 -1 1 1 1 0 0 0 0 0
0 0 0 0 0 0 1 -1 -1 -1 0 0 0 0 0
0 0 0 -1 1 0 0 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0 0 0 0 -1
0 0 -1 0 1 0 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 0 0 -1
0 0 0 0 -1 0 0 0 0 0 0 0 0 0 1
0 0 1 0 0 0 0 0 0 0 0 -1 0 0 0
0 0 1 0 0 0 0 0 0 0 0 0 -1 0 0
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1 1 1 1 1 1 1 0 0 0 0 0


KERNEL

 matrix dimension r3 x c15
-0 -0 -0 -0 -0 -0 -0 1 -1 -0 -0 -0 -0 -0 -0
-0 -0 -0 -0 -0 -0 -0 1 -0 -1 -0 -0 -0 -0 -0
-1 -1 -1 -1 -1 -1 -1 -1 -0 -0 -1 -1 -1 -1 -1

 enzymes

 1: (2) R13 -R14
 2: (2) R13 -R15
 3: (13) -R02 -R04 -R07 -R08 -R10 -R11 -R12 -R13 -R01 -R03 -R05 -R06 -R09

 overall reaction
```

```
 1: CREB(PS) + PFKFB-2 = CREB + PFKFB-2(PS)
 2: PFKFB-2 + S6(P) = PFKFB-2(PS) + S6
 3: 3 ADP + GDP + 4 NDP + 7 P + PFKFB-2(PS) = 3 ATP + GTP + 4 NTP + PFKFB-2


CONSERVATION RELATIONS

 matrix dimension r10 x c22
1 1 -0 -0 -0 -0 -0 -0 0 0 0 -0 0 -0 0 -0 -0 -0 0 0 0 0
-0 0 1 1 -0 -0 -0 -0 1 0 0 -0 0 -0 0 -0 -0 -0 0 0 0 0
-0 0 -1 -1 1 1 1 1 0 1 0 -0 0 -0 0 -0 -0 -0 0 0 0 0
-0 0 -1 -1 1 1 1 1 0 0 1 -0 0 -0 0 -0 -0 -0 0 0 0 0
-0 0 -0 -0 -0 -0 -0 -0 0 0 0 1 1 -0 0 -0 -0 -0 0 0 0 0
-0 0 -0 -0 -0 -0 -0 -0 0 0 0 -0 0 1 1 -0 -0 -0 0 0 0 0
-0 0 -0 -0 -0 -0 -0 -0 0 0 0 -0 0 -0 0 -1 -1 1 1 0 0 0
-0 0 -0 -0 -0 -0 -0 -0 0 0 0 -0 0 -0 0 1 1 -0 0 1 0 0
-0 0 -1 -1 -0 -0 1 1 0 0 0 -0 0 -0 0 -0 -0 -0 0 0 1 0
-0 0 -1 -0 -0 -0 -0 -0 0 0 0 -0 0 -0 0 -0 -0 -0 0 0 0 1

 1: ERK + ERK(P) = const
 2: Grb-2 + Grb-2_Sos + IIR(PY)_Shc(PY)_Grb-2_Sos = const
 3: -Grb-2 + -Grb-2_Sos + IIR + IIR(PY) + IIR(PY)_Shc + IIR(PY)_Shc(PY) + InsR =
const
 4: -Grb-2 + -Grb-2_Sos + IIR + IIR(PY) + IIR(PY)_Shc + IIR(PY)_Shc(PY) + Insulin =
const
 5: MEK + MEK(P) = const
 6: RSK + RSK(P) = const
 7: -Raf + -Raf(P) + Ras_GDP + Ras_GTP = const
 8: Raf + Raf(P) + Ras_GTP_Raf(P) = const
 9: -Grb-2 + -Grb-2_Sos + IIR(PY)_Shc + IIR(PY)_Shc(PY) + Shc = const
 10: -Grb-2 + Sos = const


SUBSETS

 matrix dimension r4 x c15
1 1 1 1 1 1 1 0 0 0 1 1 1 1 1
0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0

 enzymes

 1: (12) R02 R04 R07 R08 R10 R11 R12 R01 R03 R05 R06 R09 irreversible
 2: (1) R13 irreversible
 3: (1) R14 irreversible
 4: (1) R15 irreversible

 overall reaction

 1: RSK + 2 ATP + GTP + 4 NTP = RSK(P) + 2 ADP + GDP + 4 NDP + 6 P
```

```
 2: RSK(P) + ATP + PFKFB-2 = RSK + ADP + P + PFKFB-2(PS)
 3: RSK(P) + ATP + CREB = RSK + ADP + CREB(PS) + P
 4: RSK(P) + ATP + S6 = RSK + ADP + P + S6(P)


REDUCED SYSTEM

 matrix dimension r2 x c4
-1 1 1 1
1 -1 -1 -1
The following line indicates reversible (0) and irreversible reactions (1)
1 1 1 1


ELEMENTARY MODES

 matrix dimension r3 x c15
1 1 1 1 1 1 1 1 0 0 1 1 1 1 1
1 1 1 1 1 1 1 0 1 0 1 1 1 1 1
1 1 1 1 1 1 1 0 0 1 1 1 1 1 1

 enzymes

 1: (13) R02 R04 R07 R08 R10 R11 R12 R13 R01 R03 R05 R06 R09 irreversible
 2: (13) R02 R04 R07 R08 R10 R11 R12 R14 R01 R03 R05 R06 R09 irreversible
 3: (13) R02 R04 R07 R08 R10 R11 R12 R15 R01 R03 R05 R06 R09 irreversible

 overall reaction

 1: 3 ATP + GTP + 4 NTP + PFKFB-2 = 3 ADP + GDP + 4 NDP + 7 P + PFKFB-2(PS)
 2: 3 ATP + CREB + GTP + 4 NTP = 3 ADP + CREB(PS) + GDP + 4 NDP + 7 P
 3: 3 ATP + GTP + 4 NTP + S6 = 3 ADP + GDP + 4 NDP + 7 P + S6(P)
```

Beitragende Autoren

Angaben zum Eigenanteil

Titel	**Literaturangabe**	**Autoren**	**Arbeitsanteil**
Analysis of structural robustness of metabolic networks.	*IEE Proceedings - Systems Biology*, 1(1):114–120, 2004.	Thomas Wilhelm, **Jörn Behre**, Stefan Schuster	40% 35% 25%
Structural robustness of metabolic networks with respect to multiple knockouts.	*Journal of Theoretical Biology*, 252(3):433–441, 2008.	**Jörn Behre**, Thomas Wilhelm, Axel von Kamp, Eytan Ruppin, Stefan Schuster	40% 20% 15% 5% 20%
Modelling signal transduction in enzyme cascades with the concept of elementary flux modes.	*Journal of Computational Biology*, 16(6):829–844, 2009.	**Jörn Behre**, Stefan Schuster	60% 40%

....

bestätigt durch Prof. Dr. Stefan Schuster

Über den Autor

Lebenslauf

Persönliche Daten

Name: Jörn Arnold Behre
Geburtsdatum: 28. Juli 1973
Geburtsort: Bergisch Gladbach, Deutschland
Nationalität: deutsch
Familienstand: ledig

Ausbildung

seit Juni 2004	Doktorand am Lehrstuhl für Bioinformatik an der Biologisch-Pharmazeutischen Fakultät der Friedrich-Schiller-Universität in Jena, Deutschland
Juli 2004	Praktikum *Modellierung biologischer Systeme* am Lehrstuhl für theoretische Biophysik der Humboldt-Universität zu Berlin, Deutschland
März 2001 – April 2001	Praktikum in C-Programmierung am Rechenzentrum der Universität zu Köln, Deutschland
Oktober 1993 – November 2002	Studium der Chemie an der Universität zu Köln, Deutschland mit Spezialisierung auf Biochemie, Diplom in Chemie
1984 – 1993	Besuch des Nicolaus-Cusanus-Gymnasiums in Bergisch Gladbach, Deutschland, Abitur in 1993

Sprachen

- Deutsch (Muttersprache)
- Englisch (fließend)
- Französisch (Schulkenntnisse)
- Latein (großes Latinum)

Publikationen

1. **J. Behre, L.F. de Figueiredo, S. Schuster, and C. Kaleta.** *Detecting structural invariants in metabolic networks.* Buchkapitel in J. van Helden, A. Toussaint, and D. Thieffry (Eds.), Methods in Molecular Biology: Bacterial Molecular Networks (Methods and Protocols). Verlag Humana Press, Kapitel 20, ISBN 978-1-61779-360-8, 2012.
2. **T. Hinze, J. Decraene, G.G. Mitchell, J. Behre, and S. Schuster.** *Towards a Unified Approach for the Modelling, Analysis, and Simulation of Cell Signalling Networks.* Buchkapitel in Z. Zhao (Ed.), Sequence and Genome Analysis: Methods and Applications. Verlag iConcept Press, Kapitel 5, ISBN 978-0-9807330-4-4, 2010.
3. **J. Behre, G. Escuela, and T. Hinze (Eds.).** *Proceedings Compendium of the Fourth Workshop on Membrane Computing and Biologically Inspired Process Calculi (MeCBIC2010) and of the First Workshop on Applications of Membrane Computing, Concurrency, and Agent-based Modelling in Population Biology (AMCA-POP2010).* Verlag ProBusiness Berlin, ISBN 978-3-86805-767-6, 2010, 238 Seiten.
4. **C. Gille, C. Bölling, A. Hoppe, S. Bulik, S. Hoffmann, K. Hübner, A. Karlstädt, R. Ganeshan, M. König, K. Rother, M. Weidlich, J. Behre, and H.-G. Holzhütter.** *HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology.* Molecular Systems Biology, 6:411, 2010.
5. **C. Kaleta, L.F. de Figueiredo, J. Behre, and S. Schuster.** *EFMEvolver: Computing elementary flux modes in genome-scale metabolic networks.* In I. Grosse, S. Neumann, S. Posch, F. Schreiber, and P.F. Stadler (Eds.), Lecture Notes in Informatics - Proceedings of the 24th German Conference on Bioinformatics, Gesellschaft für Informatik, Bonn, 179–189, 2009.
6. **J. Behre and S. Schuster.** *Modeling signal transduction in enzyme cascades with the concept of elementary flux modes.* Journal of Computational Biology, 16(6), 829–844, 2009.

7. **J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, and S. Schuster.** *Structural robustness of metabolic networks with respect to multiple knockouts.* Journal of Theoretical Biology, 252(3), 433–441, 2008.

8. **T. Wilhelm, J. Behre, and S. Schuster.** *Analysis of structural robustness of metabolic networks.* IEE Proceedings - Systems Biology, 1(1), 114–120, 2004.

9. **J. Behre** *Analyse von Konformationsänderungen bei der Komplexbildung von Proteinen.* Diploma thesis, 2002.

Vorträge

1. **J. Behre, G. Escuela, and T. Hinze.** *Eröffnungsvortrag*[*] Eleventh International Conference on Membrane Computing (CMC11), Jena, Deutschland, 24.08.2010.

2. **J. Behre and T. Hinze.** *Eröffnungsvortrag*[*] Fourth Workshop on Membrane Computing and Biologically Inspired Process Calculi (MeCBIC), Jena, Deutschland, 23.08.2010.

3. **J. Behre and S. Schuster.** *Structural modelling of signal transduction in enzyme cascades with the concept of elementary flux modes.* International Workshop on Metabolic Pathways Analysis (MPA), Leiden, Niederlande, 27.10.2009.

4. **J. Behre and S. Schuster.** *Calculating elementary flux modes in enzyme cascades.* JCB-Seminar, Jena, Deutschland, 18.12.2008.

5. **J. Behre and S. Schuster.** *Structural Modelling of Signal Transduction in Hepatocytes exemplified by the Insulin Network.* HepatoSys Platform Meeting "Modeling", Berlin, Deutschland, 24.10.2008.

6. **J. Behre, T. Wilhelm, A. von Kamp, and S. Schuster.** *Structural robustness of metabolic networks on the basis of elementary flux modes.* 3rd International IEEE Scientific Conference on Physics and Control - PhysCon, Potsdam, Deutschland, 03.09.2007. (Eingeladener Vortrag)

7. **J. Behre.** *Comparison of networks in cell biology.* SFB 604 "Multifunctional Signaling Proteins" - 4th SFB Workshop "Gene regulatory networks", Dornburg, Deutschland, 07.12.2006. (Eingeladener Vortrag)

8. **T. Wilhelm, J. Behre, and S. Schuster.** *Analysis of structural robustness of metabolic networks on the basis of elementary flux modes.* JCB-Seminar, Jena, Deutschland, 11.11.2004.

[*] zusammen mit Kollegen

Posterpräsentationen

1. **J. Behre, R. Samaga, S. Klamt, and S. Schuster.** *Structural Modelling of Signal Transduction in Hepatocytes exemplified by the Insulin Network.* JCB Workshop II, Jena, Deutschland, 20. November 2009.

2. **J. Behre, R. Samaga, S. Klamt, and S. Schuster.** *Structural Modelling of Signal Transduction in Hepatocytes exemplified by the Insulin Network.* German Conference on Bioinformatics, Halle, Deutschland, 26.–28. September 2009.

3. **J. Behre, R. Samaga, S. Klamt, and S. Schuster.** *Structural Modelling of Signal Transduction in Hepatocytes exemplified by the Insulin Network.* German Symposium on Systems Biology, Heidelberg, Deutschland, 12.–15. Mai 2009.

4. **J. Behre, R. Samaga, S. Klamt, and S. Schuster.** *Structural Modelling of Signal Transduction in Hepatocytes exemplified by the Insulin Network.* Evaluation des HepatoSys-Projekts, Berlin, Deutschland, 21.–23. Januar 2009.

5. **J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, and S. Schuster.** *Structural robustness of metabolic networks based on the concept of elementary flux modes.* Conference on Systems Biology of Mammalian Cells (SBMC), Dresden, Deutschland, 22.–24. Mai 2008.

6. **J. Behre, T. Wilhelm, A. von Kamp, S. Klamt, and S. Schuster.** *Structural robustness of metabolic networks based on the concept of elementary flux modes.* Statusseminar des HepatoSys-Projekts, Heidelberg, Deutschland, 15.–16. Oktober 2007.

7. **J. Behre, T. Wilhelm, A. von Kamp, E. Ruppin, and S. Schuster.** *Structural robustness of metabolic networks based on the concept of elementary flux modes.* German Conference on Bioinformatics, Potsdam, Deutschland, 26.–28. September 2007. **(Best Poster Award)**

8. **J. Behre, T. Wilhelm, and S. Schuster.** *Structural robustness of metabolic networks on the basis of elementary flux modes.* Posterbeitrag für das Jena Centre for Bioinformatics (JCB) bei der Evaluierung der deutschen Bioinformatik-Zentren, Berlin, Deutschland, 01.–02. März 2007.

9. **J. Behre, T. Wilhelm, and S. Schuster.** *Structural robustness of metabolic networks on the basis of elementary flux modes.* Conference on Systems Biology of Mammalian Cells (SBMC), Heidelberg, Deutschland, 12.–14. Juli 2006.

10. **J. Behre, T. Wilhelm, and S. Schuster.** *Structural robustness of metabolic networks with respect to single and multiple knockouts on the basis of elementary flux modes.* 12th Meeting of the BTK: International Study Group for Systems Biology, "Systems Biology: redefining BioThermoKinetics", Trakai, Litauen, 14.–17. September 2006.

11. **J. Behre, T. Wilhelm, and S. Schuster.** *Robustness of metabolic networks on the basis of elementary flux modes.* Evaluierungs- & Statusseminar des HepatoSys-Projekts, Berlin, Deutschland, 28.–30. November 2005.

12. **J. Behre, T. Wilhelm, and S. Schuster.** *Robustness of metabolic networks on the basis of elementary flux modes.* European Conference on Mathematical and Theoretical Biology - ECMTB, Dresden, Deutschland, 18.–22. Juli 2005.

13. **J. Behre, T. Wilhelm, and S. Schuster.** *Calculating robustness of metabolic networks on the basis of elementary flux modes.* International Workshop on Metabolic Pathways Analysis (MPA), Jena, Deutschland, 03.–05. März 2005.

14. **J. Behre, T. Wilhelm, and S. Schuster.** *Calculating robustness of metabolic networks on the basis of elementary flux modes.* 5th International Conference on Systems Biology - ICSB, Heidelberg, Deutschland, 09.–13. Oktober 2004.

15. **J. Behre, T. Wilhelm, and S. Schuster.** *Calculating robustness of metabolic networks on the basis of elementary flux modes.* German Conference on Bioinformatics, Bielefeld, Deutschland, 04.–06. Oktober 2004.

Lehre

Übungen, Seminare und Praktika

Max-Planck-Institut für Chemische Ökologie in Jena

- Praktikum zum IMPRS (International Max Planck Research School) Graduiertenkolleg *Modelling of Metabolic Networks**, International Max Planck Research School
  (Wintersemester 2008/2009)

Friedrich-Schiller-Universität Jena

- Praktikum zur Vorlesung *Metabolische und regulatorische Netzwerke**, Studiengang Bioinformatik
  (Wintersemester 2004/2005, Wintersemester 2005/2006, Sommersemester 2007 and Wintersemester 2008/2009)
- Übungen zur Vorlesung *3D-Strukturen biologischer Makromoleküle*, Studiengang Bioinformatik
  (Wintersemester 2007/2008, Sommersemester 2006)

*Betreuung zusammen mit Kollegen

- Übungen zur Vorlesung *Optimality Principles in Evolution*, Studiengang Bioinformatik
  (Wintersemester 2006/2007)
- Proseminar *Recherche in molekularbiologischen Datenbanken**, Studiengang Bioinformatik
  (alle Sommersemester über die Jahre 2004–2009)
- Übungen zur Vorlesung *Einführung in die Bioinformatik II 1**, Studiengang Bioinformatik
  (Sommersemester 2006, Sommersemester 2009)
- Übungen zur Vorlesung *Einführung in die Bioinformatik 2b**, Studiengang Bioinformatik
  (Wintersemester 2005/2006)

Universität zu Köln

- Grundpraktikum in Biochemie*,
  Institut für Biochemie, Studiengänge in Biologie & Chemie
  (Sommersemester 2002)

Betreuung von Diplomarbeiten

- S. Schäuble. *Computer Based Analysis of Nitrogen Metabolism in Chlamydomonas reinhardtii.* 2009.*
- F. Wessely. *Theoretical Analysis of the Relationship between Metabolic Network Structure and Enzyme Gene Expression.* 2008.

Jena, den 28.03.2012

....
Jörn Arnold Behre

*Betreuung zusammen mit Kollegen

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel angefertigt habe. Mir ist die geltende Promotionsordnung bekannt und ich habe weder die Hilfe eines Promotionsberaters in Anspruch genommen, noch haben Dritte unmittelbare oder mittelbare geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die vorgelegte Dissertation wurde noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht. Weiterhin habe ich mich mit der vorliegenden Arbeit an keiner anderen Hochschule um den akademischen Grad doctor rerum naturalium (Dr. rer. nat.) beworben und weder früher noch gegenwärtig die Eröffnung eines Verfahrens zum Erwerb des o.g. akademischen Grades an einer anderen Hochschule beantragt.

Bei der Auswahl und Auswertung des Materials, sowie bei der Herstellung des Manuskripts haben mich meine Kollegen am Lehrstuhl für Bioinformatik unter der Leitung von Prof. Dr. Stefan Schuster unterstützt.

Jena, den 28.03.2012

....
Jörn Arnold Behre