

Modellbasierte Identifizierung von Interventionsstrategien für Stoffwechselnetzwerke

**Dissertation
zur Erlangung des akademischen Grades**

**Doktoringenieur
(Dr.-Ing.)**

**von Diplom-Biomathematiker Oliver Hädicke
geb. am 26.05.1980 in Bergen auf Rügen**

**genehmigt durch die Fakultät für Elektrotechnik und Informationstechnik
der Otto-von-Guericke-Universität Magdeburg**

**Gutachter:
Prof. Dr.-Ing. Achim Kienle
Prof. Dr. Stefan Schuster
Dr.-Ing. Steffen Klamt**

Promotionskolloquium am 14.11.2014

Forschungsberichte aus dem Max-Planck-Institut
für Dynamik komplexer technischer Systeme

Band 43

Oliver Hädicke

**Modellbasierte Identifizierung von Interventions-
strategien für Stoffwechselnetzwerke**

Shaker Verlag
Aachen 2015

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Zugl.: Magdeburg, Univ., Diss., 2014

Copyright Shaker Verlag 2015

Alle Rechte, auch das des auszugsweisen Nachdruckes, der auszugsweisen oder vollständigen Wiedergabe, der Speicherung in Datenverarbeitungsanlagen und der Übersetzung, vorbehalten.

Printed in Germany.

ISBN 978-3-8440-3406-6

ISSN 1439-4804

Shaker Verlag GmbH • Postfach 101818 • 52018 Aachen

Telefon: 02407 / 95 96 - 0 • Telefax: 02407 / 95 96 - 9

Internet: www.shaker.de • E-Mail: info@shaker.de

Kurzzusammenfassung

Mit Hilfe der industriellen Biotechnologie können eine Vielzahl verschiedener Substanzen durch den Einsatz geeigneter, evtl. modifizierter Mikroorganismen produziert werden. Zukünftige Herausforderungen liegen unter Anderem in der Verwendung erneuerbarer und kostengünstiger Ausgangsmaterialien sowie in der Steigerung der Effizienz der eingesetzten Mikroorganismen. Daher ist ein Ziel des Metabolic Engineering, durch gezielte Eingriffe die Stoffflüsse von Mikroorganismen oder Zellen derart umzulenken, dass ein gewünschtes Produkt vermehrt gebildet wird. Die molekularbiologischen Werkzeuge zur Umsetzung solcher Eingriffe sind Knockouts bzw. Überexpressionen von Genen. Für die systemweite Identifizierung geeigneter Interventionen sind mathematische Modellierung und Simulation metabolischer Netzwerke wichtige Werkzeuge.

Zwei neue *in silico* Methoden, CASOP und constrained Minimal Cut Sets, wurden hier entwickelt. Mit diesen können Interventionen sowohl für das Strategieziel optimaler Produktausbeuten als auch für das Ziel einer hohen Produktivität vorgeschlagen werden. Mit dem EFMSampler wurde zudem ein neuer Ansatz zur Berechnung von Elementarmoden in genomskaligen Netzwerken vorgestellt, wodurch eine Vielzahl von elementarmodenbasierten Methoden für die Analyse genomskaliger Netzwerke zugänglich wird. Des Weiteren wurde ein generischer Ansatz aufgezeigt, wie falsch-positive Resultate in metabolischen Vergleichsstudien mit grundlegenden statistischen Verfahren vermieden werden können. Abschließend wurde anhand der Überexpression der Phosphoenolpyruvat Synthetase eine der neu vorgeschlagenen Interventionen zur Produktion von Lactat bzw. Succinat experimentell getestet. Die hier vorgestellten Methoden können bei der Konstruktion neuer Produktionsstämme eingesetzt werden und unterstützen somit die Entwicklung ökologisch nachhaltiger und ökonomisch kompetitiver Prozesse.

Danksagung

Die vorliegende Dissertation entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Max-Planck-Institut für Dynamik komplexer technischer Systeme in Magdeburg.

Mein erster Dank geht an Dr. Steffen Klamt, der mein Interesse am Metabolic Engineering geweckt hat und der meine Arbeit über die gesamte Zeit der Dissertation wissenschaftlich begleitet hat. Unzählige Diskussionen und eine stets offene Tür waren unentbehrlich beim Gelingen der vorliegenden Dissertation.

Ebenso möchte ich Prof. Achim Kienle (Otto-von-Guericke-Universität Magdeburg) und Prof. Stefan Schuster (Friedrich-Schiller-Universität Jena) für die Begutachtung der Arbeit danken.

Des Weiteren möchte ich Steffi Strähler für ihren Einsatz bei der experimentellen Konstruktion des *Escherichia Coli*-Stammes KBM10111(pRR48-ppsA) und der Durchführung der zugehörigen Kultivierungsexperimente danken. Katja Bettenbrock gilt der Dank für die vielen Diskussionen und die Unterstützung dieser Experimente. Weiterer Dank geht an Verena Lohr für die experimentellen Arbeiten mit der Vogelzelllinie AGE1.CR.pIX. Die Vorarbeiten von Christoph Kaleta zur Berechnung von Elementarmoden in genomskaligen Netzwerken waren für das entsprechende Kapitel dieser Arbeit entscheidend.

Wenn die Arbeiten zum Stoffwechsel von *Rhodospirillum Rubrum* auch nicht Bestandteil der vorliegenden Arbeit wurden, so möchte ich dennoch Lisa und Anke Carius, Christiane Rudolf und Hartmut Grammel für die vielen Diskussionen bezüglich den Eigenheiten des Stoffwechsel dieser besonderen Bakterien danken.

Meinen ehemaligen Bürokollegen Hannah Sharp, Daniel Samaga, Detlev Bannasch, Axel von Kamp, Sebastian Mirschel, Katrin Kolczyk und Carsten Conradi danke ich dafür, dass ich die Zeiten im Büro immer als sehr angenehm empfunden habe. Neben all den fachlichen Hilfestellungen denke ich auch gern an manche Diskussionen, die sowohl kleinere Dinge als auch das Große Ganze betreffen, zurück. Philipp Paetzold gebührt ein zusätzlicher Dank für das Korrekturlesen der Arbeit.

Ebenso möchte ich allen weiteren ehemaligen und aktuellen Kollegen während meiner Zeit am Max Planck Institut danken, die für die sehr angenehme Arbeitsatmosphäre gesorgt haben und von denen einige im Laufe der Zeit zu Freunden geworden sind.

Ganz besonderer Dank gilt zum Schluss meiner Familie. Zum Einen meinen Eltern, die mich in die Lage verstetzt haben ein derartiges Projekt angehen und durchstehen zu können. Zudem danke ich Anika, die in den letzten Monaten oftmals auf mich verzichtet hat, und meiner Tochter Finja Lina. Ihr beide sorgt dafür, dass ich genau so gerne aus dem Institut hinaus wie hinein gehe.

Inhaltsverzeichnis

1	Einleitung	1
2	Industrielle Biotechnologie und Metabolic Engineering	5
2.1	Industrielle Biotechnologie	5
2.2	Metabolic Engineering	7
3	Biologische Grundlagen	11
3.1	Molekularbiologische Methoden	11
3.1.1	Gen-Deletionen	11
3.1.2	Überexpressionen und Genexpressions-Tuning	12
3.1.3	Protein Engineering	13
3.1.4	Adaptive Evolution von Mikroorganismen	13
3.1.5	Rationale Zufallsmethoden	14
3.2	Grundlagen <i>Escherichia coli</i>	16
3.2.1	<i>Escherichia coli</i> als Modellorganismus des Metabolic Engineering	16
3.2.2	Zentralstoffwechsel von <i>Escherichia coli</i>	17
4	Mathematische Grundlagen	21
4.1	Metabolische Reaktionsnetzwerke	21
4.1.1	Rekonstruktion stöchiometrischer Netzwerke	21
4.1.2	Stöchiometrische Stoffwechselmodelle für <i>Escherichia coli</i>	22
4.1.3	Formale Definition stöchiometrischer Netzwerke	23
4.2	Analyse stationärer Flussverteilungen	25
4.2.1	Fließgleichgewicht und stationäre Flussverteilungen	25
4.2.2	Metabolische Stoffflussanalyse	26
4.3	Constraint-basierte Modellierung	28
4.3.1	Prinzipien der constraint-basierten Modellierung	28
4.3.2	Definition des Flusskegels	29
4.3.3	Flussbilanzanalyse	30
4.3.4	Flussvariabilitätsanalyse	32
4.3.5	Anwendungen der Flussbilanzanalyse	33
4.3.6	Elementarmodenanalyse	35
4.3.7	Minimal Cut Sets	38
4.4	Modellbasierte Methoden zur Identifizierung von Interventionsstrategien	42
4.4.1	Ansätze mittels mathematischer Optimierung	42
4.4.2	Ansätze mittels Elementarmodenanalyse	48
4.4.3	Ansatz mittels Minimal Cut Sets	49
4.4.4	Ansätze mit kinetischer Modellierung	49

5	Neue Methoden zur modellbasierten Identifizierung von Interventionsstrategien	53
5.1	CASOP	53
5.1.1	Formale Definition	53
5.1.2	Erweiterungen	63
5.2	Constrained Minimal Cut Sets	65
5.2.1	Definition von constrained Minimal Cut Sets	65
5.2.2	Reformulierung anderer Ansätze	70
5.2.3	Verallgemeinerung für komplexe Interventionen	73
5.3	EFMSampler	78
5.3.1	Elementarmodenberechnung für genomskalige Netzwerke	78
5.3.2	EFMSampler zur Berechnung von Elementarmodenteilmengen	78
5.3.3	Analyse der Qualität von Elementarmodenteilmengen	81
6	Anwendungen der neuen Methoden	85
6.1	Betrachtete Produkte und bekannte Interventionsstrategien	85
6.1.1	Succinat	85
6.1.2	Lactat	89
6.1.3	Lycopin	91
6.2	Berechnung von Interventionsstrategien	94
6.2.1	Anwendung von CASOP	94
6.2.2	Anwendung der constrained Minimal Cut Sets	106
6.2.3	Anwendung des EFMSampler	116
7	Statistische Betrachtung metabolischer Vergleichsstudien	119
7.1	Motivation	119
7.2	Mathematische Grundlagen	120
7.2.1	Theorie statistischer Hypothesentests	120
7.2.2	Varianzquellen extrazellulärer Raten	121
7.3	Fallstudie zum Vergleich von Kultivierungssystemen	123
7.3.1	Experimentelle Daten	123
7.3.2	Statistische Analyse der experimentellen Daten	126
7.4	Statistische Versuchsplanung	128
8	Experimentelle Überprüfung einer ausgewählten Intervention	131
8.1	Manipulation von Co-Faktoren	131
8.2	ATP-Verbrauch zur Steigerung der Produktivität	132
8.3	Experimentelle Durchführung	133
8.3.1	Überexpression von <i>ppsA</i>	133
8.3.2	Kultivierungen	134
8.3.3	Experimentelle Ergebnisse	134
8.4	Analyse der experimentellen Ergebnisse	135
8.4.1	Bestimmung der Raten und Flussverteilungen	135
8.4.2	Auswertung der Raten und Flussverteilungen	136
9	Zusammenfassung	139

A Anhang	143
A.1 Zentralstoffwechselmodell für <i>Escherichia Coli</i>	143
A.2 p -Werte der statistischen Tests mit unvollständigen Daten	148
A.3 Experimentelle Daten	149
Literaturverzeichnis	151

Liste der verwendeten Symbole und Abkürzungen

α	Signifikanzniveau
D	Menge der Desired-Elementarmoden
ΔG	Freie Gibbs'sche Energie
E	Menge aller Elementarmoden
γ	Relativer Massenanteil des Produktes an der Biomasse
k	Gewichtungparameter der CASOP-Methode
κ	Wahrscheinlichkeit, blockierte Reaktion zu entfernen (EFMSampler Parameter)
λ	Wahrscheinlichkeit, eine Knockout-Menge zu verwerfen (EFMSampler Parameter)
μ	Wachstumsrate
N	Stöchiometrische Matrix
ω	Relevanz einer Reaktion
r	Vektor der Reaktionsraten
σ	Standardabweichung
σ^2	Varianz
$\mathbb{S}$	Szenario
$t_{X/S}$	Grenzwert für Mindest-Biomasseausbeute
T	Menge der Target-Elementarmoden
Θ	Effektmaß
$\mathbb{V}$	Artifizieller externer Container
v	Elementarmodengewicht
X	Biomasse
Y	Ausbeute
Z_1, Z_2	CASOP Reaktions-Ratings
CASOP	Computational Approach for Strain Optimization aiming at high Productivity
cMCS	constrained Minimal Cut Set
EM	Elementarmodus
FBA	Flussbilanzanalyse
FVA	Flussvariabilitätsanalyse
GS	Genomskalig
Lac	Lactat
Lyco	Lycopin
MCS	Minimal Cut Set
MFA	Metabolische Stoffflussanalyse
MILP	Gemischt-ganzzahlige Optimierung (mixed integer linear programming)
MMF	Minimale metabolische Funktionalität
STD	Standardabweichung
STR	Rührkesselreaktor (stirred tank reactor)
Succ	Succinat
ZSW	Zentralstoffwechsel

1 Einleitung

Die Menschheit ist seit Urzeiten von Energie (Brennstoff) und Kohlenstoff (Nahrung) aus Biomasse abhängig. In den letzten etwa 200 Jahren ist zudem eine immer stärkere Abhängigkeit von erdöl- und erdgasbasierter Energie und erdölbasiertem Kohlenstoff hinzugekommen. Die inhärente nicht erneuerbare Natur dieser Kohlenstoffquelle führt zu einem dringendem Bedarf an alternativen Energiequellen und Herstellungsprozessen.

Die industrielle Biotechnologie bietet eine Alternative zur petrochemischen Herstellung von Brennstoffen und anderen Substanzen. Mit Hilfe von Mikroorganismen werden Zucker und andere biologische Substrate durch Fermentation effizient in gewünschte Produkte konvertiert (Jantama et al., 2008a). Auch auf Grund ökologischer Vorteile ist es äußerst erstrebenswert, zukünftig die petrochemische Synthese vieler Substanzen durch Produktionsprozesse mit Mikroorganismen zu ersetzen. Dafür bedarf es der Konstruktion von Biokatalysatoren bzw. Zellfabriken, die schnell und effizient geeignete Substrate in die Zielprodukte konvertieren. Der Gesamtprozess muss dabei kompetitiv bzgl. der Kosten im Vergleich zur petrochemischen Herstellung sein, so wie es für das Beispiel des Polylactats detailliert diskutiert wurde (Vink et al., 2007).

Die chemische Synthese ist jedoch noch immer der dominante Herstellungsprozess, da die biotechnologische Herstellung vieler Produkte noch nicht ökonomisch wettbewerbsfähig ist. Dennoch können bereits heute viele industriell wichtige Substanzen mit Zellfabriken effizient produziert werden (Lee et al., 2012b). Die Palette erfolgreicher Anwendungen reicht von Biotreibstoffen (z.B. Ethanol, Biogas, Wasserstoff) über Plattform- und Feinchemikalien (z.B. organische Säuren, Enzyme, Vitamine) bis hin zu Biopolymeren (z.B. Polyhydroxyalkanoat, Polylactat, Poly- γ -Glutamat (Lee et al., 2012a)). Für einige andere Produkte steht die mikrobielle Herstellung an der Schwelle zur ökonomischen Nutzbarkeit (Sauer et al., 2010, 2008). Darüber hinaus verspricht die physiologische Diversität der mikrobiellen Welt die Entwicklung weiterer Anwendungen zur biotechnologischen Herstellung eines breiteren Spektrums an Substanzen sowie neuer Produkte.

Die gezielte Nutzung der mikrobiellen Produktion bestimmter Substanzen hat seine Ursprünge in der frühen Antike. Schon die Sumerer (um 3000 v. Chr.) beherrschten die Herstellung von Wein durch Hefen. Weitere (unbewusste) Anwendungen von Mikroorganismen, die seit der Antike bekannt sind, umfassen z.B. die Herstellung fermentierter Nahrungsmittel wie Joghurt, Käse, Essig und Bier. Die verwendeten Mikroorganismen wurden dabei im Laufe der Zeit durch Selektion gewünschter Eigenschaften stetig verbessert, ohne dass die zu Grunde liegenden biologischen Mechanismen vollständig erfasst wurden.

Durch die Aufklärung des biologischen Codes der DNA und der Entwicklung rekombinanter DNA-Technologien ist es seit relativ kurzer Zeit möglich, von einzelnen Proteinen über mehrere Reaktionsabfolgen bis hin zu ganzen Organismen gezielte Modifikationen

vorzunehmen, um gewünschte Eigenschaften in diese Systeme zu implementieren. Das Konzept der Zellfabriken wurde in den 1980er Jahren industriell etabliert (z.B. durch die biotechnologische Herstellung von Insulin (Johnson, 1983)). Zum Einsatz kamen dabei zunächst modifizierte Wildtyp-Stämme von z.B. *Escherichia coli* (*E. coli*) oder *Saccharomyces cerevisiae*, jedoch wurden diese bald durch spezialisierte Stämme mit deutlich verbesserten Eigenschaften ersetzt (Makino et al., 2011; Lee et al., 2005; Park et al., 2008; Lee et al., 2009; Na et al., 2010; Lee et al., 2011; Adrio und Demain, 2006, 2010; Demain und Adrio, 2008, 2012; Chen et al., 2010; Kondo et al., 2013).

Diese Verbesserungen wurden zunächst durch die Anwendung ungerichteter Mutagenese und anschließender phänotypischer Selektion erreicht. Dieser Ansatz wird nach und nach durch die Anwendung von Methoden des Metabolic Engineering, der Systembiologie und der Synthetischen Biologie ersetzt. Die Integration der Methoden dieser drei Bereiche kann als Systems Metabolic Engineering (Lee et al., 2011, 2012a; Wittmann und Lee, 2012) beschrieben werden und umfasst verschiedene methodische Werkzeuge, um metabolische Netzwerke zu optimieren.

Die Einflüsse der Fermentationsbedingungen lassen sich sehr gut empirisch untersuchen. So können zum Beispiel verschiedene Medienzusammensetzungen, Temperaturbereiche oder Sauerstoffverfügbarkeiten systematisch experimentell untersucht werden. Die Optimierung des Mikroorganismus selbst stellt sich als ungleich komplexer dar. Je nach verwendetem Stamm besitzt *E.coli* in etwa 4600 Gene. Da jedes dieser Gene durch Überexpression bzw. Deletion Ziel einer Intervention sein kann, ergeben sich bereits 9200 Einzeleingriffe. Werden zudem Interventionsstrategien mit bis zu drei Modifikationen betrachtet, so sind über $7.7 \cdot 10^{11}$ verschiedene Strategien möglich. Eine extensive experimentelle Untersuchung und Charakterisierung allein dieser begrenzten Definition von möglichen Strategien (z.B. ist eine Erweiterung auf Inhibierung statt vollständigem Knockout denkbar) ist nicht realisierbar.

Algorithmen, mit denen das gesamte Spektrum optimaler und suboptimaler Routen für die Produktsynthese analysiert werden kann, sind daher zwingend erforderlich. Für die systemweite Identifizierung geeigneter Interventionen sind mathematische Modellierung und Simulation metabolischer Netzwerke wichtige Werkzeuge. Mit Hilfe der Constraint-Based Modellierung wurden bereits viele Methoden zur Analyse von Stoffwechselnetzwerken entwickelt (Trinh et al., 2009; Lewis et al., 2012). Dabei standen meist Fragen der Strukturanalyse metabolischer Netzwerke sowie die Bestimmung optimaler Biomasse- oder Produktausbeuten im Vordergrund. In der letzten Dekade wurden neue mathematische Methoden entwickelt, die spezielle biotechnologische Fragestellungen des Metabolic Engineering adressieren (Trinh et al., 2009; Copeland et al., 2012; Zomorodi et al., 2012; Tomar und De, 2013).

Eine spezielle Klasse von Algorithmen befasst sich dabei mit der Identifizierung von Interventionsstrategien, die intrazelluläre Stoffflüsse eines Organismus so umlenken, dass eine höhere Produktausbeute erreicht wird. Der Schwerpunkt der vorliegenden Arbeit liegt auf dieser Klasse mathematischer Methoden. Dafür wird zunächst ein Überblick über existierende Ansätze gegeben sowie deren zu Grunde liegenden Ideen diskutiert. Des Weiteren werden neue Methoden vorgestellt, mit denen die Limitierungen bestehender Ansätze vermieden werden können. Die neuen Methoden werden an industriell relevanten Beispielen demonstriert, indem Interventionsstrategien für *E. coli* identifiziert werden, die zu einer gesteigerten Produktion ausgewählter Substanzen führen.

Als eine weitere Aufgabenstellung wird die Berechnung von Elementarmoden in genomskaligen Netzwerken betrachtet. Die vollständige Enumeration aller Elementarmoden ist in genomskaligen Netzwerken nicht realisierbar (Klamt und Stelling, 2003). Daher wird ein Algorithmus vorgeschlagen, der eine unverzerrte Teilmenge von Elementarmoden bestimmt, welche die wesentlichen Eigenschaften der vollständigen Menge gut erfassen kann. Vergleichende metabolische Studien können zeit- und kostenintensive Änderungen biotechnologischer Prozesse nach sich ziehen. Es ist somit von größter Wichtigkeit, dass die Ergebnisse solcher Studien statistisch belastbar sind. Daher wird ein allgemeines Vorgehen bei der Analyse solcher Studien zur Vermeidung falsch-positiver Resultate präsentiert und an einer Beispielstudie demonstriert. Abschließend wird eine der neu vorgeschlagenen Interventionen zur Produktion von Lactat bzw. Succinat experimentell getestet.

Aufbau der Arbeit

Kapitel 2 beginnt mit einem Überblick über die industrielle Biotechnologie und erläutert das Metabolic Engineering als eine zu Grunde liegende Technologie.

In **Kapitel 3** werden zunächst verfügbare molekularbiologische Methoden zur experimentellen Implementierung von Interventionen vorgestellt. Im zweiten Abschnitt dieses Kapitels wird die Verwendung von *E. coli* als Modellorganismus motiviert und die Grundlagen seines Stoffwechsels vorgestellt.

In **Kapitel 4** werden die mathematischen Grundlagen zur Analyse stationärer Flussverteilungen und die Grundprinzipien der constraint-basierten Modellierung erläutert. Insbesondere wird das Konzept der Elementarmoden und der Minimal Cut Sets vorgestellt. Des Weiteren werden existierende mathematische Ansätze zur Identifizierung von Interventionsstrategien ausführlich diskutiert.

In **Kapitel 5** werden neue Ansätze zur Identifizierung von Interventionsstrategien vorgestellt. Die erste Methode, CASOP, ist eine elementarmodenbasierte Heuristik, die es explizit ermöglicht, zwischen der Ausbeute und der Produktivität als Design-Parameter der Mutante abzuwägen. Die zweite Methode, die constrained Minimal Cut Sets, sind eine Verallgemeinerung der Minimal Cut Sets und erleichtern die Formulierung komplexer Interventionsprobleme. Zudem lassen sich viele andere Methoden als spezielle constrained Minimal Cut Set Probleme reformulieren. Abschließend wird in diesem Kapitel der EFMSampler zur Berechnung von Elementarmoden-Teilmengen in großen Netzwerken vorgestellt und die Qualität der generierten Teilmengen eingehend untersucht.

Kapitel 6 demonstriert die neuen Methoden anhand geeigneter Fallstudien. Dabei werden für Produkte des Zentral- und Sekundärstoffwechsels genetische Interventionsstrategien für *E. coli* vorgeschlagen, die zu einer gesteigerten Synthese der betrachteten Produkte führen können.

In **Kapitel 7** wird die Vorgehensweise bei der statistischen Bewertung experimenteller Daten von metabolischen Vergleichsstudien diskutiert. Dies wird an einer Beispielstudie demonstriert, bei der die Auswirkungen der Änderung des Kultivierungssystems auf die Wachstumsparameter einer Zelllinie untersucht werden.

Abschließend wird in **Kapitel 8** eine ausgewählte Intervention experimentell getestet. Insbesondere wird der Effekt eines künstlich erhöhten ATP-Verbrauchs auf die Ausbeute und Produktivität für Lactat und Succinat diskutiert.

Wesentliche Ergebnisse der vorliegenden Arbeit wurden bereits vorab veröffentlicht. So wurde die CASOP-Methodik in (Hädicke und Klamt, 2010) vorgestellt und die constrained Minimal Cut Sets in (Hädicke und Klamt, 2011) eingeführt. Die statistische Betrachtung von vergleichenden metabolischen Studien wurde zudem in (Hädicke et al., 2013) behandelt. In einem Buchbeitrag zu stöchiometrischen Netzwerkanalysen wurden die grundlegenden Konzepte und einige der hier diskutierten Methoden erläutert (Klamt et al., 2014). Auf eine explizite ständige Zitierung dieser Referenzen wird in der hier vorliegenden Arbeit verzichtet.

Es entstanden während der Dissertation weitere thematisch verwandte Publikationen, die hier aber nicht behandelt werden können. Dies betrifft folgende Arbeiten:

- In (Hädicke et al., 2011) wurden mit Hilfe von Flussvariabilitätsanalyse unter anderem Interventionsziele zur Produktion von Wasserstoff in schwefelfreien Purpurbakterien identifiziert.
- Das Konzept der constrained Minimal Cut Sets wurde in (Gruchattka et al., 2013) angewendet, um Interventionsziele für die Produktion von Terpenoiden mit *E. coli* zu identifizieren.
- Die Flussvariabilitätsanalyse wurde ebenfalls in (Carius et al., 2013) verwendet, um den metabolischen Zustand von *Rhodospirillum rubrum* unter mikroaeroben Zuständen zu beschreiben.
- Schließlich wurden Ergebnisse zu umfangreichen Flussanalysen des Stoffwechsels in einer tierischen Produktionszelle in (Lohr et al., 2014) publiziert.

2 Industrielle Biotechnologie und Metabolic Engineering

In diesem Kapitel werden grundlegende Konzepte der industriellen Biotechnologie und des Metabolic Engineering erläutert. Die Biotechnologie wird oftmals in drei Bereiche unterteilt. Einer davon ist die industrielle Biotechnologie, die auch als „Weiße Biotechnologie“ bezeichnet wird. Sie beschäftigt sich mit biotechnologischen Methoden zur Herstellung von z.B. Grund- oder Feinchemikalien als auch zur Energiegewinnung. Sie steht in Abgrenzung zur roten und grünen Biotechnologie, die auf medizinische Fragestellungen bzw. Anwendungen auf genetisch modifizierte Nutzpflanzen fokussiert sind (Soetaert und Vandamme, 2010).

2.1 Industrielle Biotechnologie

Ein großer Vorteil der weißen Biotechnologie ist die Verwendung nachwachsender Rohstoffe, was enorme ökologische Vorteile gegenüber der petrochemischen Herstellung bedeutet. Darüber hinaus ergeben sich oftmals weitere Vorteile wie deutlich weniger Abfallprodukte, ein reduzierter Energiebedarf und ein reduzierter Chemikalieneinsatz.

Dennoch sind es meist nicht die ökologischen Vorteile, die zu einer Umstellung eines chemischen hin zu einem biotechnologischen Herstellungsprozess führen. Die entscheidenden Kriterien sind vielmehr Verbesserungen der Qualitätsmaße eines Prozesses wie zum Beispiel eine höhere Produktionsrate oder Produktausbeute, eine höhere Produktreinheit sowie eine Kostenreduzierung. Die ökologischen Vorteile sind meist lediglich ein erfreulicher Nebeneffekt und für sich selbst genommen nicht hinreichend, um Entscheidungsträger zu motivieren, einen etablierten Prozess durch eine neue Technologie zu ersetzen (Soetaert und Vandamme, 2010).

In einer OECD-Studie aus dem Jahr 2001 wurden die ökologischen und ökonomischen Vorteile der industriellen Biotechnologie anhand von 21 Fallstudien demonstriert. Eine ausführliche Potentialanalyse der industriellen Biotechnologie für den Standort Deutschland wurde kürzlich vom Fraunhofer-Institut für System- und Innovationsforschung vorgestellt (Nusser et al., 2007). Die Anwendungsmöglichkeiten der industriellen Biotechnologie sind extrem vielseitig, wie die folgende (unvollständige) Auswahl an erfolgreichen Beispielen belegt:

- Chemische Industrie
 - Biotreibstoffe (z.B. Ethanol, Wasserstoff)
 - Plattformchemikalien (z.B. Succinat, Lactat)
 - Biokunststoffe (z.B. Polyhydroxybutyrat (PHB), Polylactat (PLA))
- Pharmazeutische Industrie
 - Antibiotika (z.B. Cephalosporin)
 - Hormone (z.B. Insulin, Cortison)
 - Impfstoffe (z.B. Influenza)
- Lebensmittelindustrie
 - Aminosäuren (z.B. Lysin, Glutamat)
 - Vitamine (z.B. Riboflavin (Vitamin B2))
 - Enzyme (z.B. Lipasen, Proteasen)

Die erhöhte Effizienz biotechnologischer Prozesse führt grundsätzlich zu deutlich reduzierten Umweltbelastungen. Um gleichzeitig die Wettbewerbsfähigkeit gegenüber chemischen Syntheseprozessen zu wahren, werden folgende Anforderungen an die verwendeten Produktionsstämme gestellt (Stephanopoulos und Stafford, 2002):

1. Kultivierbar mit hohen Zelldichten unter Produktionsbedingungen
2. Wachstum mit (preis-)günstigen Substraten
3. Keine Ausscheidung von Nebenprodukten (für hohe Ausbeuten und geringe Aufreinigungskosten)
4. Hohe Produktausscheideraten (Produktivität)

Um die geforderten Eigenschaften zu erreichen wurden relevante Organismen lange Zeit durch ungerichtete Methoden modifiziert. So wurde z.B. eine erhöhte Mutagenese durch Zugabe bestimmter Chemikalien oder Strahlung induziert. Anschließend wurden die Mutanten mit den gewünschten Charakteristika selektiert. Welche Mutationen dabei schließlich zu einer Verbesserung geführt haben, blieb oft unbekannt. Ein weiterer Nachteil dieser Herangehensweise ist, dass so erzeugte Produktionsstämme später nur schwer an veränderte Umweltbedingungen angepasst werden können.

Die Verfügbarkeit neuer molekularbiologischer Techniken erlaubt es inzwischen, gezielte Eingriffe in den Stoffwechsel der Zellen vorzunehmen. Dadurch ist es möglich, die Nachteile der ungerichteten Stammentwicklung zu umgehen und Produktionsstämme gezielt zu konstruieren.

2.2 Metabolic Engineering

Metabolic Engineering ist der Teilbereich der industriellen Biotechnologie, der sich mit Methoden zur Optimierung genetischer und regulatorischer Prozesse innerhalb von Zellen befasst, um bestimmte Substanzen im industriellen Maßstab ökologisch nachhaltig und ökonomisch effizient zu produzieren. Als eigenständiges Forschungsgebiet ist es durch seinen Schwerpunkt auf Integration und systemweite Analyse biosynthetischer Pfade innerhalb des Stoffwechsels charakterisiert (Reed et al., 2010).

Im engeren Sinn beschäftigt sich Metabolic Engineering mit den Prozessen *innerhalb* der Zelle. Die generelle Strategie des Metabolic Engineering umfasst darüber hinaus weitere Fragestellungen. Der gesamte Arbeitsablauf von der Zielstellung bis hin zur Etablierung eines Produktionsstammes kann in folgende Bereiche unterteilt werden (Lee et al., 2009):

1. Wahl eines geeigneten Ausgangsstammes
2. Modellgestützte (*in silico*) Prädiktion von Interventionsstrategien (siehe Kapitel 4.4 und 5)
3. Experimentelle Implementation der genetischen Manipulationen (siehe Kapitel 3)
4. Betrachtung des Gesamtprozesses (z.B. Bestimmung relevanter Prozessparameter wie Ausbeute, Produktivität, End-Titer (Zhuang et al., 2013))
5. Iteration: evtl. weitere Modifikationen am Stamm oder Prozess

Der Fokus dieser Arbeit liegt auf dem zweiten Schritt, also der Beschreibung theoretischer Methoden auf der Basis stöchiometrischer Modelle. Ziel dieser Methoden ist die Prädiktion von Interventionsstrategien zur Optimierung bestimmter Eigenschaften von Produktionsstämmen. Daher werden im Folgenden zunächst die grundlegenden Ziele dieser Interventionen erläutert.

Prinzipien von Interventionsstrategien

Die Projektion aller stationären Flussverteilungen in die zweidimensionale Biomasse- vs. Produktebene beschreibt die Phänotypebene und bildet alle theoretisch möglichen Phänotypausprägungen (bezüglich dieser beiden Merkmale) eines Organismus ab (Abbildung 2.1). Für Bakterien konnte gezeigt werden, dass sie (unter substratlimitierten Bedingungen) optimal in Bezug auf Biomasseausbeute wachsen (Edwards et al., 2001). Der konkrete Phänotyp dieser Organismen ist somit durch einen Punkt der Phänotypebene gegeben, an dem die maximale Biomasse erreicht wird und (außer evtl. essentiellen Nebenprodukten) kein weiteres Produkt ausgeschieden wird.

Sofern Wachstums- und Produktionsphase gleichzeitig ablaufen, ist der Bereich gewünschter Phänotypen meistens durch die blaue Ellipse in Abbildung 2.1 gekennzeichnet. In diesem Bereich zeigt der Organismus ein ausreichendes Wachstum und scheidet zusätzlich das gewünschte Produkt mit einer hohen Ausbeute aus. Das Ziel von genetischen Interventionen ist deshalb, die Stoffflüsse der Organismen so umzulenken, dass die resultierenden Phänotypen sich in diesem Bereich ausprägen. Dabei können zwei unterschiedliche Herangehensweisen verfolgt werden:

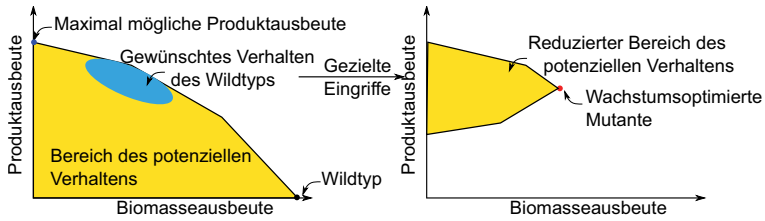


Abbildung 2.1: Zielstellung des Metabolic Engineering. Links: Phänotypenebene des Wildtyps. Der blau gekennzeichnete Bereich beschreibt die gewünschten Phänotypen. Rechts: reduzierter Bereich zulässiger Phänotypen einer Mutante mit obligatorischer Produktausscheidung.

Lokale und/oder direkte Eingriffe

Ein offensichtlicher Ansatz ist die Verstärkung oder Abschwächung von unmittelbar an der Produktsynthese beteiligten Reaktionsflüssen. Die Überexpression von Genen, die für die Synthese oder den Export des gewünschten Produktes verantwortlich sind, ist ein Beispiel für Eingriffe dieser Kategorie. Als direkte Knockout-Kandidaten bieten sich Gene an, die Enzyme kodieren welche wiederum für Pfade von Nebenprodukten wichtig sind. Diese lokalen und offensichtlichen Eingriffe führen im Allgemeinen jedoch meistens nicht dazu, dass der zulässige Bereich potentieller Phänotypen reduziert wird. Es ist somit nicht ausgeschlossen, dass der Phänotyp der Mutante zwar zunächst in den gewünschten Bereich verschoben wird, im Laufe der Zeit jedoch (zum Beispiel unter Selektionsdruck) zum Verhalten des Wildtyps zurückkehrt.

Systemweite Eingriffe

Ein komplexerer bzw. differenzierterer Ansatz ist die systemweite Identifizierung geeigneter Eingriffe. Ziel dieser Interventionen kann zum Beispiel sein, die Phänotypenebene derart zu reduzieren, dass sie in etwa dem zuvor definierten Bereich entspricht. Dies hätte zur Folge, dass die zulässigen Phänotypen der Mutante sich ausschließlich im gewünschten Bereich manifestieren können. Zudem kann (bei entsprechendem Design) garantiert werden, dass die Flussverteilung der Mutanten nicht zu der Ausgangsverteilung des Wildtyps zurückkehren kann. Vielmehr kann die Phänotypenebene so gestaltet werden, dass (unter der Annahme, dass die Mutanten über mehrere Generationen hinweg erneut optimales Wachstumsverhalten anstreben) die Produktsynthese im Laufe der Zeit von den Organismen weiter verbessert wird (Abb. 2.1). Dieses Verhalten ist Grundlage eines zentralen Konzeptes für die Identifizierung von systemweiten Interventionen, das im Folgenden erläutert wird.

Kopplung von Biomasse- und Produktsynthese

Der Ansatz der Kopplung von Biomasse- und Produktsynthese ist von entscheidender Bedeutung bei der Identifizierung von Interventionsstrategien. Die Kopplung von Produktsynthese und Zellwachstum hat zwei entscheidende Vorteile. Zum Einen kann eine bestimmte minimale Ausbeute garantiert werden und zum Anderen kann die Produktivität durch den gezielten Einsatz von adaptiver Evolution verbessert werden (siehe Kapitel 3.1.4).

Anhand des Beispielnetzwerkes in Abbildung 2.2 soll dieses Konzept erläutert werden. Das Ziel sei die Optimierung der Produktion des Metaboliten P aus S, wofür zwei potentielle Knockout-Strategien betrachtet werden. Strategie 1 besteht aus den Deletionen der Reaktionen R2 und R4, so dass suboptimale Pathways bzgl. der Ausbeute für Biomasse und Produkt blockiert werden. Die zweite Strategie besteht aus den Deletionen der Reaktionen R3 und R7.

Nach den Knockouts der Strategie 1 verbleiben drei Pathways im Netzwerk, die entweder Biomasse oder das Produkt P mit maximaler Ausbeute synthetisieren. Da die Biomasse- und Produktpfade jedoch unabhängig voneinander operieren, kann die Mutante ihren Stoffwechsel derart gestalten, dass ausschließlich Biomasse aus dem aufgenommenen Substrat gebildet wird. Somit würde im ungünstigsten Fall überhaupt kein Produkt ausgeschieden.

Die Strategie 2 hingegen hat zur Folge, dass, sobald die Mutante Biomasse synthetisiert, auch deren Präkursoren D und E gebildet werden müssen. Letzterer kann nur über Reaktion 4 (und weiter R5 und R6) gebildet werden, bei der auch P synthetisiert wird. Da es keinen alternativen Abbauweg für P gibt, muss es ausgeschieden werden, um eine ausgeglichene Flussverteilung aufrecht erhalten zu können. Somit ist mit dieser Strategie die Produktausscheidung obligatorisch an die Biomassesynthese gekoppelt.

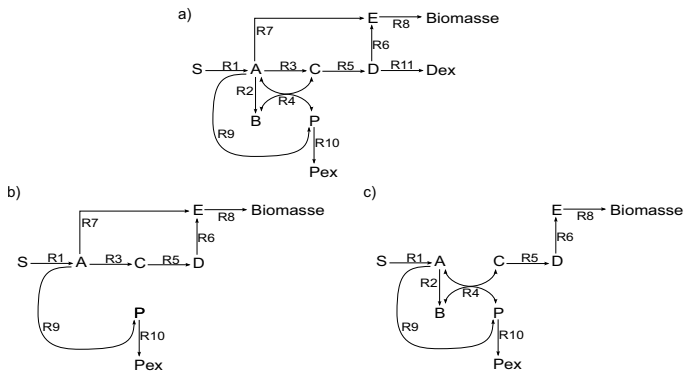


Abbildung 2.2: Kopplung von Produkt- und Biomassesynthese. a) Reaktionsnetzwerk ohne Interventionen. b) Netzwerk nach Deletion der Reaktionen R2 und R4. Diese entfernt zwar den ausbeutesuboptimalen Pfad zur Synthese von P, bewirkt aber die Entkopplung der Produkt- von der Biomassesynthese. c) Eine alternative Strategie ist die Deletion von R3 und R7, die zur obligatorischen Ausscheidung von P führt, wenn Biomasse gebildet wird.

In realen biochemischen Netzwerken wird die Ausscheidung eines Produktes an das Wachstum oftmals mittels Co-Faktoren gekoppelt. So kann zum Beispiel für Mikroorganismen durch geeignete Interventionen die Ausscheidung einer bestimmten organischen Säure unter anaeroben Bedingungen als essenzielle NADH-Senke implementiert werden. Diese Säuren (von denen einige industriell relevante Produkte darstellen) müssen also obligatorisch ausgeschieden werden, um überschüssiges NADH zu regenerieren und so ihren Redoxhaushalt auszugleichen.

3 Biologische Grundlagen

In Kapitel 3.1 wird zunächst ein Überblick über molekularbiologische Werkzeuge zur experimentellen Implementierung genetischer Interventionen gegeben. Anschließend wird in 3.2 *Escherichia coli* als Modellorganismus vorgestellt und die Grundlagen seines Stoffwechsels erläutert.

3.1 Molekularbiologische Methoden

Intrazelluläre Flüsse sind nicht unmittelbar experimentell steuerbar. Vielmehr können die Reaktionsraten nur über indirekte Methoden beeinflusst werden. Dies kann auf der Ebene der Enzyme, der RNA oder der DNA geschehen. Die entsprechenden molekularbiologischen Werkzeuge (um gezielte Interventionen in Organismen vorzunehmen) haben sich in den vergangenen Jahren rasant entwickelt. Im Folgenden wird ein kurzer Überblick über verfügbare biologische Methoden gegeben, mit denen es möglich ist, intrazelluläre Flussverteilungen zu manipulieren und somit auch modellgetriebene Hypothesen zu validieren.

3.1.1 Gen-Deletionen

Deletionsmutanten können durch homologe Rekombination erzeugt werden, wodurch das Zielgen durch eine beliebige (meist mit einer spezifischen Antibiotikaresistenz kodierenden) DNA-Sequenz ersetzt wird. Diese kann in einem nachfolgendem Schritt wieder aus dem Genom entfernt werden. Zwei verschiedene Ansätze stehen dafür zur Verfügung, die sich sowohl zur Gendeletion als auch zur Genintegration eignen. Der erste Ansatz verwendet Plasmide, die einen wählbaren Marker enthalten, flankiert von (zum Zielgen) homologen Sequenzen. Mit Hilfe von Restriktionsendonukleasen wird die DNA-Sequenz an vordefinierten Stellen geschnitten und die neu entstandenen DNA-Fragmente können dann (unter Umständen mit weiteren Fragmenten) mit Hilfe von Ligasen neu verknüpft werden (Martinez-Morales et al., 1999). Voraussetzung für neue Verbindungen ist das Vorliegen von sticky- bzw. blunt ends (komplementäre DNA-Abschnitte mit bzw. ohne Überhang). Der zweite Ansatz erlaubt eine direkte Disruption der DNA durch lineare DNA-Fragmente mittels des FRT/FLP-Systems (Datsenko und Wanner, 2000). Auch bei dieser Methode wird das chromosomale Gen durch einen wählbaren Marker ersetzt, der später entfernt werden kann (Pósfai et al., 1997). Das FRT/FLP-System verwendet die Rekombinase Flippase (Flp) des Bakteriophagen λ , um DNA-Sequenzen basengenau auszutauschen. Die Flippase bindet an beide Enden der FRT-Sequenz und rekombiniert sie.

Nach der Rekombination verbleiben jedoch Teile der FRT-Sequenzen auf der chromosomalen DNA, was die Effizienz weiterer Gendeletion reduzieren kann (Martinez-Morales et al., 1999). Dieser Ansatz wurde gewählt, um die Keio-Kollektion (Baba et al., 2006) aufzubauen; eine Sammlung von 3985 Mutanten des *E. coli*-Stammes K-12, bei denen jeweils genau ein Gen deletiert wurde.

3.1.2 Überexpressionen und Genexpressions-Tuning

Die Erhöhung eines bestimmten Flusses einer Reaktion kann durch Überexpression der zugehörigen Gene adressiert werden. Unter Überexpression versteht man die deutlich über der Norm liegende Expression eines bestimmten Gens. Daraus folgt in der Regel eine verstärkte Synthese des Proteins, für das das Gen kodiert und in der Folge ein erhöhter Fluss der Reaktion (die durch das Enzym katalysiert wird). Für eine Überexpression stehen verschiedene molekularbiologische Methoden zur Verfügung.

So kann z.B. die Expressionsrate eines Gens durch die Verwendung artifiziieller Promotoren beliebig moduliert werden (Hammer et al., 2006; Alper et al., 2005a; Solem und Jensen, 2002). Mit diesem Ansatz kann nicht nur eine starke Überexpression, sondern auch eine Feinjustierung der Genexpression erreicht werden. Eine graduelle Abstufung der Promotorstärke wird dafür durch Sequenzvariationen zwischen den konservierten -10- und -35-Bereichen bakterieller Promotoren bewirkt (Jensen und Hammer, 1998a,b). Durch den Einsatz von degenerierten Primern lassen sich unterschiedlich starke Promotoren herstellen, die den ursprünglichen Zielpromotor ersetzen. Mit dieser Methodik konnte die Expression des glykolytischen *pgi*-Gens (Glukosephosphat-Isomerase) in *E. coli* graduell um das Fünffache verstärkt bzw. abgeschwächt werden (Braatsch et al., 2008).

Alternativ können multiple chromosomale Kopien des Gens erzeugt werden oder der Anteil des entsprechenden Gens durch Plasmidlokalisation des Gens erhöht werden. In Prokaryoten werden Überexpressionen am häufigsten durch die Verwendung von Plasmidvektoren realisiert. Dabei werden die zu exprimierenden Gene auf einem Plasmid eingebaut und anschließend in den Organismus transformiert, wo es über den Kultivierungszeitraum hinweg stabilisiert werden muss. Mit dieser Methode lässt sich eine hohe Gendosis erreichen, die über die Plasmidkopienzahl gesteuert werden kann. Auch eine Modifikation der zugehörigen mRNA zur Erhöhung der Haltbarkeit ist eine mögliche Methode.

Die gesteigerte Expression kann sowohl kontinuierlich oder als induzierbar implementiert werden. Bei der ersten Methode wird das Gen permanent über einen konstitutiven Promoter exprimiert (womöglich mit intrazellulärer Akkumulation des zugehörigen Proteins). Bei der induzierbaren Expression kann durch Zugabe eines Aktivators (wie z.B. IPTG) die Stärke des Promotors beeinflusst werden, und damit auch der Expression der Gene, die unter dessen Kontrolle stehen.

3.1.3 Protein Engineering

Neben dem Redesign von Mikroorganismen können auch auf der Ebene der Enzyme Manipulationen vorgenommen werden. Durch gezielte Aminosäuresubstitutionen oder Ansätze über zufällige Substitutionen (durch Mutationen) können bestimmte Eigenschaften der Proteine implementiert bzw. verbessert werden.

Durch gerichtete Evolution von Proteinen können bestimmte Eigenschaften von Enzymen sehr schnell optimiert werden, selbst wenn keine strukturellen oder mechanistischen Informationen zur Verfügung stehen (Petrounia und Arnold, 2000). Dafür wird mit Hilfe zufälliger Mutagenese (Petrounia und Arnold, 2000), Rekombination eines Zielgens (Zhao et al., 1998) oder einer Gruppe zugehöriger Gene (Cramer et al., 1998) eine Protein-Bibliothek erstellt. Die erzeugten Proteinvarianten werden anschließend mit Hochdurchsatzmethoden bezüglich gewünschter Eigenschaften analysiert. Mit dieser Vorgehensweise können Enzymeigenschaften wie Aktivität, Löslichkeit, Selektivität und Stabilität deutlich verbessert werden (Wang et al., 2000; Cirino und Arnold, 2002).

Durch Protein Engineering können nicht nur bestehende Charakteristika von Enzymen verbessert, sondern auch neue enzymatische Aktivitäten implementiert werden. Wenn genügend Informationen über die Struktur und Funktion der Enzyme zur Verfügung stehen, können deren Eigenschaften auch gezielt durch rationale Ansätze modifiziert werden (Leisola und Turunen, 2007; Korkegian et al., 2005). Mit Hilfe spezieller Analysen ist es möglich, vorherzusagen, welche Änderungen der Aminosäuresequenz vorgenommen werden müssen, um die gewünschten Eigenschaften zu erhalten (Leisola und Turunen, 2007; Korkegian et al., 2005).

3.1.4 Adaptive Evolution von Mikroorganismen

Die zuvor beschriebenen Methoden dienen zumeist der gezielten Manipulation einzelner Reaktionsraten. Ein komplementärer Ansatz, um bestimmte (komplexe) Eigenschaften (wie z.B. Produktivität oder Robustheit) der Mikroorganismen zu verbessern, ist durch die Ausnutzung adaptiver Evolutionsprozesse gegeben.

Adaptive Evolution ist klar von nicht-adaptiver Evolution abzugrenzen. Das klassische Beispiel der *nicht-adaptiven Evolution* ist genetische Drift, d.h. dass die Frequenz von bestimmten Allelen in der Population (und mit ihnen vererbare Merkmale) sich allein auf Grund stochastischer Ursachen ändert. Ein weiteres Beispiel (in der Tierwelt) ist die sexuelle Selektion, die zwar zu Vorteilen beim Fortpflanzungserfolg führen kann, jedoch keine direkte Anpassung an bestimmte Umweltbedingungen darstellt und sich sogar als nachteilig erweisen kann (z.B. locken auffällige Farben und Töne Fressfeinde an).

Adaptive Evolution oder auch natürliche Selektion beschreibt die Fähigkeit eines Organismus, sich an bestimmte Umweltbedingungen anzupassen. Es ist ein kontinuierlicher und dynamischer Prozess, der die physische Fitness unter gegebenen Umweltbedingungen erhöht. Der vorsätzliche Einsatz dieser Methode mit biotechnologischem Hintergrund wird oftmals auch als *Evolutionary Engineering* bezeichnet (Patnaik, 2008; Sauer, 2001).

Voraussetzung des Evolutionary Engineering ist ein geeignetes Selektionskriterium. Um die Produktivität von Mikroorganismen zu erhöhen, ist dieses Kriterium durch die Wachstumsrate gegeben. Verhalten sich die ersten Generationen von Zellen nach Interventionen (Änderung der Umweltbedingungen oder Eingriffe in das Genom) oftmals suboptimal im Hinblick auf ihr Wachstumsverhalten, so adaptieren die nachfolgenden Ge-

nerationen ihren Stoffwechsel hin zu optimalen (höheren) Wachstumsraten (Ibarra et al., 2002). *E. coli* erreicht dies meist durch eine Erhöhung der Kapazität bereits aktiver Pathways und nicht durch die Aktivierung von bis dahin inaktivierten Routen (Fong et al., 2006). Sind nach der Implementierung genetischer Interventionen Wachstum und Produktausscheidung gekoppelt, so werden durch die Selektion schnell wachsender Mikroorganismen (durch serielle Transfers) gleichzeitig Zellen mit hoher Produktivität koselektiert.

Diese Methodik ist eine exzellente Ergänzung der rationalen Ansätze zur Stammoptimierung und wurde bereits mehrfach erfolgreich als Werkzeug des Metabolic Engineering eingesetzt. So wurden zum Beispiel Lactat- (Fong et al., 2005), Succinat- (Jantama et al., 2008a) oder Lycopin- (Alper und Stephanopoulos, 2007) produzierende *E. coli* Stämme mit diesem Ansatz optimiert.

3.1.5 Rationale Zufallsmethoden

Der Ansatz der gTME-(global Transcription Machinery Engineering) Methodik ist es, die Expressionslevel einzelner oder multipler Gene durch Manipulation der globalen Transkriptionsfaktoren zu perturbieren und Veränderungen selbst in komplexen Phänotypen zu induzieren (Alper und Stephanopoulos, 2007). Mit dieser Methodik konnte zum Beispiel sowohl die Lycopin-Produktion als auch die Ethanolresistenz von *E. coli* verbessert werden (Alper et al., 2006; Alper und Stephanopoulos, 2007). Das Whole Genome Shuffling (WGS) ist eine weitere Methode, mit der innerhalb kurzer Zeit gewünschte Phänotypen generiert werden können (Zhang et al., 2002; Patnaik et al., 2002). Es ermöglicht die gerichtete Evolution auf der Ebene der Mikroorganismen durch multiple rekursive Rekombinationen des gesamten Genoms. Die technischen Einzelheiten sowie ein Überblick der erfolgreichen Anwendungen werden in (Gong et al., 2009) und (Biot-Pelletier und Martin, 2014) diskutiert. Eine weitere Methode, mit der die Expression mehrerer Gene simultan beeinflusst werden kann, basiert auf der Erstellung von Bibliotheken von „Tunable Intergenic Regions“ (TIGRs). Sie wurde anhand der Überproduktion von Mevalonat mit *E. coli* demonstriert, bei der eine Erhöhung der Produktion um das Siebenfache erreicht werden konnte (Pfleger et al., 2006).

Daneben existieren weitere Methoden wie SCALE (Lynch et al., 2007), MAGE (Wang et al., 2009) oder TRMR (Warner et al., 2010), die hier jedoch nicht weiter diskutiert werden können. Alle hier erwähnten sonstigen Methoden gehören in die Kategorie der „rationalen Zufallsmethoden“, da die Maßnahmen, die zu den gewünschten Effekten führen, *a priori* nicht bekannt sind und die Mutanten nachträglich selektiert werden müssen. Weitere Vor- und Nachteile von einigen der hier vorgestellten (und weiterer) Methoden zur Genmanipulation werden zum Beispiel in den Reviews von (Warner et al., 2009) und (Patnaik, 2008) diskutiert.

Die Kombination der verfügbaren Methoden des Metabolic Engineering ergibt ein mächtiges Werkzeug für das gezielte Stammdesign (Abb. 3.1). Es gibt bereits viele erfolgreiche Beispiele (siehe oben) ihrer Anwendung, mit denen biologische Systeme nach unseren Vorstellungen umgestaltet und für die Produktion von Chemikalien, Nahrung oder Medikamenten eingesetzt werden. Dennoch ist unsere Fähigkeit noch immer recht limitiert, vorhergesagtes Verhalten biologischer Systeme schnell und verlässlich experimentell zu überprüfen bzw. zu validieren (Endy, 2005). Um dieses Ziel zu erreichen, sind daher weitere neue biologische und theoretische Methoden unerlässlich.

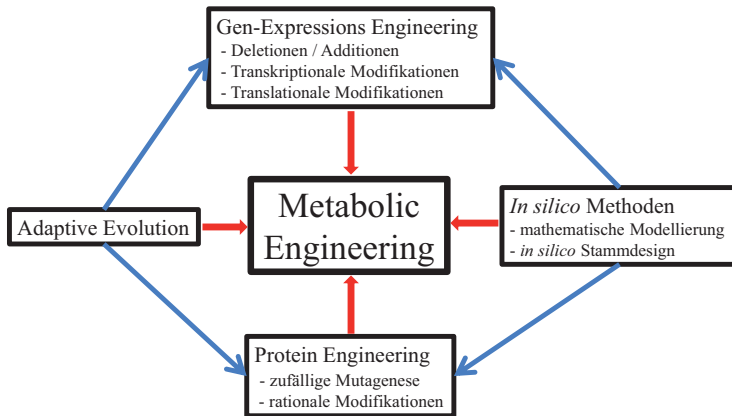


Abbildung 3.1: Werkzeuge des Metabolic Engineering

3.2 Grundlagen *Escherichia coli*

Escherichia coli ist ein gramnegatives, fakultativ anaerobes Bakterium, das im Darm von Säugetieren lebt. Es ist eines der am besten untersuchten prokaryotischen Modellorganismen, da es sich sehr einfach und kostengünstig im Labor kultivieren lässt. In der Wissenschaft kommt es hauptsächlich in den Gebieten der Biotechnologie als Produktionsstamm und in der Mikrobiologie für einen Großteil der Arbeiten mit rekombinanter DNA zur Anwendung.

3.2.1 *Escherichia coli* als Modellorganismus des Metabolic Engineering

Unter den Kandidaten für Zellfabriken bzw. Biokatalysatoren ist *E. coli* einer der prominentesten Vertreter und wurde bereits erfolgreich in vielen verschiedenen biotechnologischen Verfahren verwendet. Besonders als Produzent von rekombinanten Proteinen oder organischen Säuren hat *E. coli* mehrere Vorteile (Yu et al., 2011):

- Viele genetische Werkzeuge wurden für *E. coli* entwickelt.
- Sensitivität gegenüber vielen Antibiotika (hilfreich zur Selektion).
- Sehr gute Informationslage bezüglich des Genoms und Proteoms.
- Sehr umfangreiches Wissen über den zentralen Kohlenstoffwechsel.
- Schnelles Wachstum in Minimalmedien (mit C5- oder C6-Zucker).
- Hohe Toleranz gegenüber organischen Säuren.

Diese Vorteile machen *E. coli* zu einem universellem Produktionsorganismus, mit dem durch geeignete Interventionen eine Vielzahl von Stoffen synthetisiert werden können. Einige Beispiele umfassen die Produktion von Succinat (Sánchez et al., 2005), Malat (Jantama et al., 2008a), Lactat (Zhou et al., 2005), Ethanol (Trinh et al., 2008), L-Alanin (Zhang et al., 2007), L-Tyrosine (Chávez-Béjar et al., 2008), L-phenylalanine (Báez-Viveros et al., 2007), L-Threonin (Lee et al., 2007), Lycopin (Alper et al., 2005b) und Polylactide (PLA) (Jung et al., 2010). Selbst heterologe Stoffwechselwege können in *E. coli* mit vertretbarem Aufwand exprimiert werden, so dass die möglichen Anwendungen auch Substanzen umfassen, für die im Wildtyp keine Synthesewege vorhanden sind. Allein die Anwendungen zur Produktion von Medikamenten (bzw. deren Wirkstoffen) sind extrem vielseitig und reichen von Anti-Oxidantien (Lycopin), Anti-Tumoren (Echinomycin), antiviralen Medikamenten (Shikimisäure), Malariawirkstoffen (Artemisinin), Hormonen (z.B. HGH) bis hin zu Immunologischen Agenzien (z.B. IFN γ) (Lee et al., 2009). Ein ausführlicher Überblick etablierter Interventionen und ihrer Implikationen für den Zentralstoffwechsel von *E. coli* wird im Review von (Papagianni, 2012) gegeben.

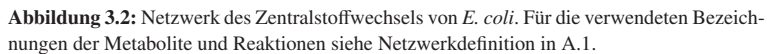
3.2.2 Zentralstoffwechsel von *Escherichia coli*

Die Kohlenstoffaufnahme aller Organismen wird durch ein integriertes Netzwerk der Signaltransduktion, der Genexpression sowie des Metabolismus gesteuert. Signaltransduktionsnetzwerke beschreiben die Erfassung und Weitergabe von Informationen zwischen Molekülen, während Genregulationsnetzwerke die regulatorischen Wechselwirkungen zwischen Genen und Proteinen beschreiben. Metabolische Netzwerke sind Abfolgen biochemischer (Enzym-katalisierter) Reaktionen, die die Umsetzung von Ausgangsmolekülen (Substraten bzw. Nährstoffen) zu definierten Endmolekülen (Produkte, Biomasse) ermöglichen. Durch die hohe Komplexität der biologischen Vorgänge lassen sich viele Phänomene nur schwer intuitiv erklären. Obwohl für kaum einen anderen Organismus ein so tiefes Verständnis des Stoffwechsels vorliegt wie für *E. coli*, ist es kaum möglich, eine umfassende mathematische Beschreibung des gesamten Stoffwechsels mit allen Interaktionen zwischen den Netzwerkklassen vorzunehmen. Daher wird der Zentralstoffwechsel von *E. coli* häufig durch die isolierte Betrachtung des metabolischen Netzwerkes vereinfacht beschrieben.

E. coli kann eine Vielzahl verschiedener Substrate nutzen, darunter verschiedene C5- und C6-Zucker sowie Glycerol und einige organische Säuren. Seine benötigte Energie kann es sowohl mittels Respiration als auch durch gemischte Säuregärung gewinnen (fakultativ anaerob). Bei Letzterem wird hauptsächlich Lactat, Ethanol und Acetat ausgeschieden (in kleineren Mengen finden sich auch Formiat, Succinat und CO₂). Die grundlegenden katabolischen Stoffwechselwege für die Metabolisierung von Glukose sind in Abbildung 3.2 skizziert. Zu den wichtigsten Pfaden gehören unter anderem

- die Glykolyse und die Glukoneogenese,
- der Entner-Doudoroff-Weg,
- der Pentose-Phosphat-Weg,
- der Citratsäurezyklus (bzw. Krebs-Zyklus oder TCA),
- der Glyoxylat-Shunt,
- die anaplerotischen Reaktionen
- sowie die Fermentationswege.

Der anabolische Stoffwechsel ist deutlich weniger verzweigt, dafür mit zum Teil komplexeren und längeren Reaktionswegen versehen. Primär fließen die Präkursoren linear zur Bildung der Makromoleküle ab. Die Makromoleküle wiederum werden in einem bestimmten Verhältnis zur Biomasse zusammengesetzt (Abbildung 3.3).



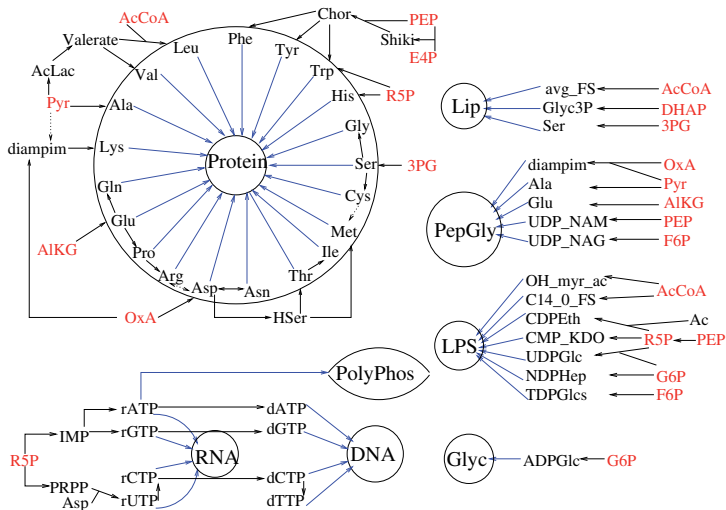


Abbildung 3.3: Synthese der Makromoleküle der Biomasse. Für die verwendeten Bezeichnungen der Metabolite und Reaktionen siehe Netzwerkdefinition in A.1.

Verstoffwechselung von Glukose in *Escherichia coli*

Der Glukosemetabolismus startet mit dessen Aufnahme zum Beispiel mittels des Zucker-PTS-Systems. Nach der Aufnahme der Glukose kann sie via Glykolyse bis zum Pyruvat umgewandelt werden. Mittels der Pyruvat-Dehydrogenase wird dieses unter aeroben Bedingungen zu Acetyl-CoA und CO_2 oxidiert. Unter anaeroben Bedingungen wird ein Teil des Pyruvats als Lactat ausgeschieden. Acetyl-CoA verknüpft die Glykolyse mit dem TCA, dem Glyoxylat-Shunt und dem Acetat-Stoffwechsel. Es kann entweder im TCA vollständig zu CO_2 oxidiert werden, in der Fettsäuresynthese zur Synthese von Biomasse verwendet werden oder im Glyoxylat-Shunt zu Malat umgesetzt werden. Unter anaeroben Bedingungen kann es zudem als Acetat oder über die Acetaldehyd-Dehydrogenase als Ethanol ausgeschieden werden. Die Regulation des Acetyl-CoA-Pools ist somit einer der Schlüsselfaktoren für die Gesamtflusslage der Zellen und beeinflusst neben der Biomassensynthese den Redox-Haushalt, den ATP-Haushalt sowie die Funktion weiterer Prozesse. Die Verknüpfung zwischen Glykolyse (bzw. Glukoneogenese bei einigen anderen Substraten) und TCA ist durch den PEP-Pyruvate-Oxalacetat-Knoten gegeben. Phosphoenolpyruvat (PEP) und Pyruvat können entweder über Acetyl-CoA in den TCA geschleust werden oder über die anaplerotischen Reaktionen die Intermediate Malat und Oxalacetat des TCA regenerieren. Unter Verwendung spezieller Substrate, wie z.B. organischer Säuren, werden Oxalacetat und Malat zu PEP und Pyruvat decarboxyliert und generieren so die Präkursoren der Glukoneogenese. Im oxidativen Teil des Pentosephosphat-Weges (via Glukose-6-Phosphat-Dehydrogenase) wird ein Großteil des für die Biomassensynthese benötigten NADPH generiert. Der nicht-oxidative Teil umfasst die Umsetzung von Ribulose-5-Phosphat (RL5P) mit Hilfe der Transaldolase und der Transketolase zum Glyceraldehyd-3-Phosphat (G3P).

4 Mathematische Grundlagen

In diesem Kapitel wird zunächst in 4.1 die mathematische Beschreibung stöchiometrischer Netzwerke erläutert. Anschließend werden in 4.2 die grundlegenden Methoden für die Analyse stationärer Flussverteilungen vorgestellt. In 4.3 werden dann sowohl das Konzept als auch verschiedene Methoden der Constraint-Based Modellierung erläutert. Abschließend werden in 4.4 verschiedene mathematische Methoden diskutiert, mit denen Interventionsstrategien *in silico* identifiziert werden können.

4.1 Metabolische Reaktionsnetzwerke

Metabolische Reaktionsnetzwerke bestehen aus Metaboliten sowie den Stoffwechselreaktionen zwischen ihnen. Fast alle Reaktionen des Stoffwechsels werden durch Enzyme katalysiert. Die Verbindungen zwischen Reaktionen und Enzymen sind nicht zwingend eindeutig. Zum Einen können multifunktionale Enzyme mehrere verschiedene Reaktionen katalysieren. Zum Anderen kann eine Reaktion durch verschiedene Enzyme (Isoenzyme) katalysiert werden. Des Weiteren kann ein Enzym durch ein oder mehrere verschiedene Gene kodiert sein. Enzymkomplexe sind oftmals durch verschiedene Untereinheiten zusammengesetzt, die wiederum durch verschiedene Gene kodiert sind. Um organismus-spezifische metabolische Netzwerke zu konstruieren, ist es demnach nötig, Informationen über dessen Gene sowie den Verknüpfungen zwischen Genen, Enzymen und Reaktionen zu integrieren (Feist et al., 2009; Thiele und Palsson, 2010).

4.1.1 Rekonstruktion stöchiometrischer Netzwerke

Für die Rekonstruktion metabolischer Netzwerke stehen neben der Primärliteratur und Übersichtsartikeln weitere Datenquellen und Hilfsmittel zur Verfügung. Umfangreiche Datenbanken wie z.B. KEGG (Kanehisa et al., 2014; Kanehisa, 1997), EcoCyC (Keseler et al., 2013) oder BioCyC (Caspi et al., 2010) dokumentieren Informationen über vorhandene Gene und Reaktionen für eine Vielzahl von Organismen. Des Weiteren existiert eine Reihe von spezialisierter Software, um den Prozess der Netzwerkrekonstruktion zu unterstützen. Eines dieser Programme, Model SEED (Henry et al., 2010), unterstützt die automatische Rekonstruktion genomweiter Netzwerke, indem es die vollständige Genomsequenz eines Organismus nach annotierten Genen durchsucht und schließlich das potentiell verfügbare Stoffwechselnetz automatisiert und standardisiert generiert.

Dennoch ist es oftmals nötig, spezielle manuell kuratierte Netzwerke zu erstellen. Häufig handelt es sich um Netzwerke des Zentralstoffwechsels (in der Größenordnung von bis zu etwa 300 Reaktionen), um fokussierte Analysen durchzuführen, für die kein genomskaliges Netzwerk benötigt wird. Zudem sind einige Analysen wie zum Beispiel die vollständige Enumeration aller Elementarmoden auf Grund ihrer kombinatorischen Komplexität in genomskaligen Netzwerken nicht realisierbar.

4.1.2 Stöchiometrische Stoffwechselmodelle für *Escherichia coli*

Genomskalige Netzwerke

Für den häufigsten Laborstamm von *E. coli*, K-12 MG1655, existieren die wohl vollständigsten und detailliertesten modellbasierten Rekonstruktionen des Stoffwechsels eines Organismus überhaupt. Das erste genomskalige stöchiometrische Modell für *E. coli* (iJE660) wurde von Edwards erarbeitet (Edwards und Palsson, 2000b). Es war das bis dato umfangreichste Stoffwechselmodell, dass jemals konstruiert wurde. In den folgenden Jahren gab es mehrfach neue Versionen und Erweiterungen dieses Netzwerkes. Das erste Update, iJR904 (Reed et al., 2003), führte zu einer Erhöhung der Anzahl der betrachteten Metabolite und Reaktionen (z.B. wurden Pathways für die Aufnahme alternativer Substrate inkludiert). Zusätzlich wurden Gen-Protein-Reaktions-Assoziationen (GPRs) eingeführt und alle Reaktionen wurden auf Konsistenz bezüglich sämtlicher beteiligter Elemente und Ladungen geprüft. In der folgenden Aktualisierung, iAF1260 (Feist et al., 2007), wurde erneut der Umfang erweitert (z.B. wurden viele Reaktionen für die Synthese von Zellwandkomponenten aufgenommen) und alle Metabolite wurden einem der Kompartimente Cytoplasma, Periplasma oder extrazellulärer Raum zugewiesen. Zudem wurden thermodynamische Eigenschaften aller Reaktionen berechnet und genutzt, um untere Schranken für (vorhergesagte) irreversible Reaktionen zu definieren. Dieses Modell (iAF1260) enthielt bereits 2077 Reaktionen, 1039 Metabolite und 1260 Gene. Das letzte Update, iJO1366 (Orth et al., 2011), wurde erneut um einige weitere charakterisierte Gene und Reaktionen erweitert. Die Änderungen bzw. Erweiterungen fallen jedoch deutlich kleiner aus, da das iAF1260 Modell bereits ein sehr umfassendes Abbild des Stoffwechsels darstellt. Dennoch wurden auf Grund neuer Erkenntnisse weitere Gene und Reaktionen hinzugefügt, so dass iJO1366 die zur Zeit vollständigste Rekonstruktion des Stoffwechsels von *E. coli* ist. Alle zuvor genannten Modelle enthalten zudem eine kumulierte Biomassesyntesereaktion, die den Bedarf der Präkursoren und Co-Faktoren (wie z.B. ATP) für die Bildung von Biomasse beschreibt.

Modelle des Zentralstoffwechsels

Neben den genomskaligen stöchiometrischen Modellen existieren viele weitere Netzwerkmodelle, die sich auf den Zentralstoffwechsel von *E. coli* fokussieren. Der Umfang dieser Modelle ist deutlich kleiner; sie enthalten meist nicht mehr als 100 Metabolite oder 200 Reaktionen. Für viele Fragestellungen ist der Detailgrad dieser Modelle dennoch ausreichend. Auch für die Identifizierung von Interventionsstrategien haben sich diese Netzwerkmodelle als sehr erfolgreich erwiesen. So wurden unter anderem Knockout-Targets für die Produktion von Ethanol, Carotinoiden und effiziente Biomassesyntese *in silico* bestimmt und experimentell validiert (Trinh et al., 2006; Trinh und Sreenc, 2009; Trinh et al., 2008; Unrean et al., 2010).

In der vorliegenden Arbeit wird das Zentralstoffwechselmodell aus (Hädicke und Klamt, 2010) verwendet. Dieses Modell wurde für die späteren Fallstudien mit den Reaktionen des DXP (Deoxy Xylulose Phosphate) Pathway und der Lycopin-Synthese erweitert, um die Produktion von Lycopin betrachten zu können. Mit diesen umfasst es 106 Metabolite und 120 Reaktionen. Die wesentlichen katabolen Stoffwechselwege dieses Modells und die Abflüsse der Präkursoren in die Makromoleküle sind in den Abbildungen 3.2 und 3.3 dargestellt. Die vollständige Netzwerkdefinition ist im Anhang A.1 gegeben.

4.1.3 Formale Definition stöchiometrischer Netzwerke

Unabhängig von der Größe des betrachteten Netzwerkes ist eine einzelne isoliert betrachtete biochemische Reaktion wie folgt charakterisiert:

- **Stöchiometrie:** Die Stöchiometrie der Reaktion wird in der Reaktionsgleichung angegeben und spiegelt die molaren Verhältnisse der teilnehmenden Edukte (negative Koeffizienten) und Produkte (positive Koeffizienten) wider. Falls eine Reaktion reversibel ist (siehe unten), müssen Hin- und Rückrichtung spezifiziert werden, um die stöchiometrischen Koeffizienten mit Bezug zur Vorwärtsrichtung zuweisen zu können.
- **Reversibilität:** Grundsätzlich sind alle chemischen Reaktionen reversibel, da zu einem gewissen Anteil auch immer ihre Rückreaktion abläuft. Faktisch können einige biochemische Reaktionen jedoch als irreversibel angesehen werden, da sie unter biologisch relevanten Bedingungen fast ausschließlich in eine Richtung ablaufen.
- **Gen-Enzym-Reaktions-Assoziationen:** Die meisten biochemischen Reaktionen werden durch Enzyme katalysiert. Die Verknüpfungen zwischen Reaktionen, Enzymen und Genen sind oftmals nicht eindeutig. So kann eine Reaktion z.B. durch verschiedene (Iso-)Enzyme katalysiert werden. Ebenso können multifunktionale Enzyme mehrere verschiedene Reaktionen katalysieren. Zudem ist jedes Enzym einem oder mehreren (im Fall von Enzymkomplexen mit mehreren Untereinheiten) kodierenden Genen assoziiert. Die resultierenden Gen-Enzym-Reaktions-Assoziationen ermöglichen es somit genomische Informationen mit Eigenschaften des Reaktionsnetzwerkes zu verknüpfen. Zudem sind diese Informationen oft die wesentliche Quelle für die Rekonstruktion organismusspezifischer metabolischer Netzwerke (Feist et al., 2009; Thiele und Palsson, 2010).
- **Reaktionskinetiken:** Reaktionskinetiken beschreiben die Dynamik der Reaktionen basierend auf ihrem Reaktionsmechanismus (z.B. Massenwirkungs- oder Michaelis-Menten-Kinetik), den Enzymeigenschaften und den Konzentrationen der beteiligten Metabolite. In vielen Fällen sind diese detaillierten Eigenschaften unbekannt.

Stöchiometrische Netzwerkanalysen basieren hauptsächlich auf den statischen Eigenschaften des Netzwerkes und lassen die kinetischen Eigenschaften unberücksichtigt. Als Ausnahme sei die Verwendung thermodynamischer Informationen für eine bestimmte Klasse von Analysen erwähnt. So ist es z.B. möglich, mit Hilfe der Änderung der freien

Gibbs'schen Energie unter Standardbedingungen Rückschlüsse auf die Reaktionsrichtungen zu ziehen oder untere bzw. obere Grenzen für bestimmte Reaktionsraten für physiologische Konzentrationsbereiche zu bestimmen (Hoppe et al., 2007; Schellenberger et al., 2011). Wenn alle Komponenten eines metabolischen Netzwerkes zusammengestellt wurden, kann die Netzwerkstruktur formal wie folgt beschrieben werden:

- m : Anzahl der Metabolite (Spezies)
- q : Anzahl der Reaktionen
- $\mathbf{N}$: $m \times q$ stöchiometrische Matrix – jede Zeile entspricht einer Spezies und jede Spalte einer Reaktion; das Matricelement n_{ij} enthält den stöchiometrischen Koeffizienten des Metaboliten i in Reaktion j
- Rev : die Indexmenge reversibler Reaktionen
- $Irrev$: die Indexmenge der irreversiblen Reaktionen ($Rev \cap Irrev = \emptyset$)

Optional können Gen-Enzym-Reaktions-Verknüpfungen als Boolesche Funktionen für jede Reaktion berücksichtigt werden (Thiele und Palsson, 2010). Obwohl sie nicht durch Enzyme katalysiert werden, bietet es sich an, Transportprozesse (z. B. Substrataufnahme oder Austausch von Metaboliten zwischen verschiedenen Kompartimenten) und Biomasse-Synthese direkt in diesen Formalismus zu integrieren, indem man sie als Pseudo-Reaktionen auffasst. Die Biomasse-Synthesereaktion beschreibt die molaren Anforderungen von Energie (ATP), Redoxkraft (z.B. NADPH) und Monomerbausteinen (Aminosäuren, Fettsäuren, Nukleotiden etc.), um die Makromoleküle (z.B. Proteine, DNA, RNA, Lipide, etc.) zu synthetisieren, die für ein Gramm Biomasse-Trockengewicht benötigt werden.

Die Grenzen von metabolischen Stoffwechselmodellen zu ihrer umgebenden Umwelt werden durch Definition der *internen* und *externen* Metaboliten gezogen. *Interne* Spezies werden explizit im Modell bilanziert und sind daher in der stöchiometrischen Matrix $\mathbf{N}$ enthalten. Im Gegensatz dazu liegen die *externen* Spezies typischerweise außerhalb des Systems (beispielsweise Substrate oder Produkte) und können als Senken oder Quellen betrachtet werden.

Die stöchiometrische Matrix $\mathbf{N}$ ist nicht nur für die stationäre Modellierung von Stoffwechselmodellen von Bedeutung, sondern auch für die dynamische Modellierung wesentlich. Die Konzentrationsänderungen der Metabolite über die Zeit können durch das folgende Differentialgleichungssystem beschrieben werden:

$$\frac{d\mathbf{c}(t)}{dt} = \mathbf{N}\mathbf{r}(t) . \quad (4.1)$$

Der $m \times 1$ -Vektor $\mathbf{c}(t)$ enthält die Metabolitkonzentrationen, typischerweise in mmol pro Gramm Zelltrockengewicht: mmol/gDW . Der $q \times 1$ -Vektor $\mathbf{r}(t)$ beschreibt die (Netto-) Reaktionsrate zur Zeit t , in der Regel in der Einheit $\text{mmol}/(\text{gDW} \cdot h)$. Der Vektor $\mathbf{r}(t)$ wird auch als Flussvektor oder Flussverteilung bezeichnet und ist im Allgemeinen abhängig von den Metabolitkonzentrationen und mehreren Parametern $\mathbf{p}$:

$$\mathbf{r}(t) = \mathbf{f}(\mathbf{c}(t), \mathbf{p}) . \quad (4.2)$$

Wie zuvor bereits erläutert, liegen die Unsicherheiten bei der dynamischen Modellierung metabolischer Netzwerke in der exakten Formulierung von $\mathbf{f}$ und der Kenntnis von $\mathbf{p}$. Hingegen ist die strukturelle Invariante des Differentialgleichungssystems (4.1), $\mathbf{N}$, meist sehr gut bekannt. Bei vollständiger Informationslage ist auf Grund des höheren explorativen und prädiktiven Potentials die dynamische Modellierung der rein stöchiometrisch basierten vorzuziehen. Die in $\mathbf{N}$ abgebildeten strukturellen Zusammenhänge legen dem Differentialgleichungssystem jedoch bereits deutliche Einschränkungen auf. Dies erklärt die erstaunlich vielfältigen Möglichkeiten, Aussagen über Netzwerkeigenschaften zu treffen, basierend auf ausschließlich stöchiometrischen Analysen.

Strukturelle Analysemethoden, die ausschließlich auf graphentheoretischen Ansätzen beruhen, sollen im Folgenden nicht betrachtet werden. Vielmehr liegt der Fokus auf Methoden zur Analyse stationärer Flussverteilungen, für die weitere Nebenbedingungen und auch experimentelle Daten integriert werden können.

4.2 Analyse stationärer Flussverteilungen

Kausale Zusammenhänge im Stoffwechsel, wie zum Beispiel der Ausgleich des Redoxhaushalts oder der ATP-Bedarf für die Biomassesyntese, schränken die zulässigen Flussverteilungen oftmals erheblich ein. Um solche Zusammenhänge zu erkennen und aufzuklären, kann auf bewährte Methoden der Stoffflussanalyse zurückgegriffen werden. In diesem Abschnitt werden die mathematischen Grundlagen dieser Methoden erläutert und mögliche Anwendungen diskutiert.

4.2.1 Fließgleichgewicht und stationäre Flussverteilungen

Die Reaktionsgeschwindigkeiten bzw. Stoffumsätze des Zellstoffwechsels laufen im Vergleich zu regulatorischen Prozessen sehr schnell ab. Daher wird für die Analyse von Stoffwechselnetzwerken häufig die Annahme postuliert, dass die Reaktionsgeschwindigkeiten als quasistationär betrachtet und somit als konstant angesehen werden können. Diese Annahme führt unter Berücksichtigung von $\frac{dc(t)}{dt} = \mathbf{0}$ (Gleichung 4.1) zur zentralen Bilanzgleichung aller stöchiometrischen Stoffflussanalysen:

$$\mathbf{N}\mathbf{r} = \mathbf{0}. \quad (4.3)$$

Gleichung 4.3 beschreibt ein homogenes lineares Gleichungssystem. Es besagt, dass Produktion und Verbrauch eines jeden internen Metaboliten ausgeglichen sind. Dadurch, dass die Reaktionsraten ebenfalls als unveränderlich betrachtet werden, ändert sich auch die mathematische Behandlung der Reaktionsraten $\mathbf{r}$. Diese hängen nun nicht mehr von den Konzentrationen und Parametern ab (4.2), sondern werden als unabhängige Variable $\mathbf{r}$ betrachtet. Dadurch werden alle Flussvektoren $\mathbf{r}$, die Gleichung 4.3 lösen, als potentielle stationäre Flussverteilungen für die Zelle betrachtet. In dem zugehörigen Differentialgleichungssystem ist mit einem spezifischem Parametersatz normalerweise nur eine Teilmenge dieser Lösungen realisierbar. Dennoch ist die Analyse des vollständigen Lösungsraumes von Gleichung 4.3 angebracht, da er alle Bereiche potentieller Flussverteilungen (für alle möglichen Parametersätze) abdeckt (Machado et al., 2012a).

Da die Anzahl der Reaktionen q in realen Netzwerken gewöhnlich größer als die Anzahl interner Metabolite m ist, lösen unendlich viele Flussverteilungen $\mathbf{r}$ das lineare Gleichungssystem 4.3. Die Menge aller Lösungen sind im rechten Nullraum (oder dem Kern) der stöchiometrischen Matrix $\mathbf{N}$ enthalten. Die Dimension des Nullraums entspricht der Anzahl der linear unabhängigen Lösungen für Gleichung 4.3. Es können somit $q - \text{rang}(\mathbf{N})$ Basisvektoren des Nullraums berechnet und als die Spalten einer zugehörigen Kernmatrix $\mathbf{K}$ angeordnet werden. Alle anderen zulässigen Lösungen für Gleichung 4.3 können dann als Linearkombination der Spalten von $\mathbf{K}$ konstruiert werden.

Die triviale Lösung $\mathbf{r} = \mathbf{0}$ erfüllt immer Gleichung 4.3, ist aber nicht von Interesse, da dies dem thermodynamischen Gleichgewicht einer toten Zelle entspricht. Lebende Zellen hingegen müssen Fließgleichgewichte erfüllen, die sie innerhalb bestimmter Freiheitsgrade realisieren können.

4.2.2 Metabolische Stoffflussanalyse

Das Ziel der metabolischen Stoffflussanalyse (MFA) ist es, konkrete Flussverteilungen eines metabolischen Netzwerkes zu bestimmen. Gleichung (4.3) ist im Allgemeinen unterbestimmt, wobei die Dimension des Nullraumes die Anzahl an Freiheitsgraden widerspiegelt. Daher benötigt man gemessene Flusswerte, um einige bzw. sogar alle unbekannten Flüsse berechnen zu können. Intrazelluläre Flüsse können in der Regel nicht direkt experimentell gemessen werden und sind der Menge der zu berechnenden Flüsse zuzuordnen. Es ist jedoch im Regelfall möglich, mehrere Aufnahme- (wie z.B. von Kohlenstoffsubstraten oder Sauerstoff) und Ausscheideraten (z.B. CO_2 oder Fermentationsprodukte) experimentell zu messen. Zudem kann die Wachstumsrate μ (in der Regel in $[h^{-1}]$ angegeben) der Biomasse experimentell gut gemessen werden.

Um die vorgenannten Raten im Fließgleichgewicht messen zu können, werden die Mikroorganismen oder Zellkulturen unter kontrollierten Bedingungen entweder in einem Bioreaktor im kontinuierlichen bzw. stationären Betrieb oder mit einer Batch-Kultivierung unter pseudostationären Bedingungen kultiviert. Mit den gemessenen Raten kann Gleichung (4.3) (nach einer eventuellen Umordnung der Spalten in $\mathbf{N}$ und der Elemente in $\mathbf{r}$) in einen bekannten bzw. gemessenen (Index k) und einen unbekannten (Index u) Teil getrennt werden:

$$\mathbf{N}\mathbf{r} = \mathbf{N}_u\mathbf{r}_u + \mathbf{N}_k\mathbf{r}_k = \mathbf{0}. \quad (4.4)$$

Dies führt zu der zentralen Gleichung der Stoffflussanalyse:

$$\mathbf{N}_u\mathbf{r}_u = -\mathbf{N}_k\mathbf{r}_k, \quad (4.5)$$

wobei mit l die Anzahl der gemessenen Raten (in $\mathbf{r}_k$) gegeben ist und die Anzahl der unbekannten Raten in $\mathbf{r}_u$ sich durch $q - l$ errechnet. Da $\mathbf{N}_k$ und $\mathbf{r}_k$ bekannt sind, erzeugt ihr Produkt einen Vektor. Damit bildet Gleichung (4.5) ein inhomogenes lineares Gleichungssystem. Dessen allgemeine Lösung für $\mathbf{r}_u$ ist gegeben durch (Klamt et al., 2002):

$$\mathbf{r}_u = -\mathbf{N}_u^\# \mathbf{N}_k \mathbf{r}_k + \mathbf{K}_u \mathbf{a}. \quad (4.6)$$

$\mathbf{N}_u^\#$ ist dabei die Penrose-Pseudoinverse von $\mathbf{N}_u$. Sie hat die Dimension $l \times m$, existiert für jede beliebige Matrix $\mathbf{N}_u$ und mit ihrer Hilfe wird eine Kleinste-Quadrate-Lösung für Gleichung (4.6) bestimmt. $\mathbf{K}_u$ ist die Kernmatrix von $\mathbf{N}_u$ welche die Basisvektoren des Lösungsraums des homogenen Gleichungssystems $\mathbf{N}_u \mathbf{r}_u = \mathbf{0}$ enthält. Linearkombinationen aus den Spalten von $\mathbf{K}_u$ (gebildet mit $\mathbf{a}$) charakterisieren daher die Freiheitsgrade für $\mathbf{r}_u$. Im einfachsten Fall ist $\mathbf{N}_u$ eine $m \times m$ quadratische Matrix mit vollem Rang, wobei $\mathbf{N}_u^\#$ dann der Standardinversen $\mathbf{N}_u^{-1}$ entspricht und somit $\mathbf{K}_u$ leer ist. In diesem Fall ist eine eindeutige Lösung für alle unbekannten Raten berechenbar.

Im Allgemeinen kann das Gleichungssystem bzgl. Determiniertheit und Redundanz in vier Kategorien klassifiziert werden, was jeweils verschiedene Konsequenzen für die Berechenbarkeit der unbekannten Raten nach sich zieht (Klamt et al., 2002).

MFA als deskriptiver Ansatz zur Charakterisierung metabolischer Zustände

Die quantitative Beschreibung intrazellulärer Flussverteilungen mit Hilfe der MFA ist ein wichtiger Baustein bei der Identifizierung von Engpässen in metabolen Pfaden („pathway bottlenecks“) oder bei der Analyse der Netzwerkregulation (Nielsen, 2007). Zudem ist das detaillierte Wissen um die Kohlenstoffflüsse innerhalb eines betrachteten Organismus eine natürliche Voraussetzung, wenn diese gezielt umgelenkt werden sollen.

Gemessene extrazelluläre Raten können explizit für die Schätzung der intrazellulären Raten verwendet werden. Durch eine Kleinste-Quadrate-Regression werden dabei die Abweichungen zwischen experimentellen Daten auf der einen Seite und simulierten Daten auf der anderen Seite minimiert (Klamt et al., 2002). Liegen nicht genug extrazelluläre Flussdaten vor, um eine eindeutige Lösung zu bestimmen, können weitere experimentelle Daten aus ^{13}C Markierungsexperimenten integriert werden (Wiechert und de Graaf, 1996; Wiechert et al., 2001; Wiechert, 2001). Bei diesen Experimenten wird ein isotopisch markiertes Substrat verwendet und die resultierenden Markierungsmuster der intrazellulären Metabolite gemessen. Anschließend werden die Flüsse so geschätzt, dass sowohl die Abweichungen zu den gemessenen extrazellulären Raten als auch die Abweichungen zu den intrazellulären Markierungsmustern minimal werden (Wiechert et al., 1997).

4.3 Constraint-basierte Modellierung

Wie im vorhergehenden Abschnitt bereits erläutert, führt die Bedingung des Fließgleichgewichts dazu, dass die zulässigen Flussverteilungen im Nullraum der stöchiometrischen Matrix N liegen. Die zu Grunde liegende Idee des constraint-basierten Modellierungsansatzes ist zusätzliche Nebenbedingungen zu verwenden, die den zulässigen Lösungsraum weiter einschränken (Lewis et al., 2012; Price et al., 2003, 2004). Diese Nebenbedingungen können durch Informationen über physiologische, biologische oder thermodynamische Zwänge begründet sein.

4.3.1 Prinzipien der constraint-basierten Modellierung

Mathematisch werden die Nebenbedingungen als lineare Gleichungs- oder Ungleichungsnebenbedingungen formuliert. Dabei können sie in folgende drei Kategorien eingeteilt werden:

C1) Fließgleichgewicht: $Nr = 0$

C2) Kapazität/Reversibilität: $\alpha_i \leq r_i \leq \beta_i$

Üblicherweise sind untere oder obere Schranken für die Austauschflüsse (Aufnahme/Ausscheidung) bekannt. Für interne Reaktionen können v_{max} -Werte eventuell aus vorhergehenden biochemischen Studien zur Verfügung stehen. Für irreversible Reaktionen kann zudem meist $\alpha_i = 0$ gesetzt werden. Wenn keine Daten für die maximalen Reaktionsraten vorliegen, so können die Schranken dennoch entweder auf einen beliebig großen (faktisch nicht limitierenden) Wert gesetzt oder als unbeschränkt definiert werden ($\pm\infty$). Sind ausschließlich die Reversibilitätsbedingungen von Interesse, können diese Nebenbedingungen zu C2' vereinfacht werden.

C2') Reversibilität: $r_i \geq 0 \quad \forall i \in Irrev$

C3) Messungen: $r_i = b_i$ (für alle Reaktionen i mit gemessenen bzw. bekannten Raten)

Durch diese Definition wird die klassische MFA zu einem Spezialfall der constraint-basierten Methode, bei dem nur die Bedingungen C1 und C3 betrachtet werden. Dadurch wird die Limitierung des MFA-Ansatzes deutlich. Indem die Reversibilitätsbedingungen nicht berücksichtigt werden, kann es passieren, dass per MFA bestimmte Flusslagen unzulässig sind, sobald Nebenbedingungen der Kategorie C2 erfüllt werden müssen. Der klassische MFA-Ansatz ist dennoch nützlich. Sollte durch die Nebenbedingungen und/oder Messungen ein inkonsistentes (Un-) Gleichungssystem aufgestellt worden sein, liefert nur der MFA-Ansatz trotzdem eine Lösung mittels Flusssschätzung durch die Kleinst-Quadrat-Methode. Alle constraint-basierten Methoden berücksichtigen die Nebenbedingungen der Klasse C2 explizit und liefern in solchen Fällen keine Lösungen, da keine Flussverteilung existiert, die C1 bis C3 erfüllen kann.

4.3.2 Definition des Flusskegels

Falls nur die Bedingungen C1 und C2' definiert sind, bildet die Menge $\mathcal{F}$ aller Flussvektoren $\mathbf{r}$, die diesen Bedingungen genügen, einen *konvexen polyedrischen Kegel* (Rockafellar, 1970; Bertsimas und Tsitsiklis, 1997), der in stöchiometrischen Studien oftmals auch als *Flusskegel* bezeichnet wird.

$$\mathcal{F} = \{\mathbf{r} \in \mathbb{R}^q | \mathbf{N}\mathbf{r} = \mathbf{0}, r_i \geq 0 \forall i \in Irrev\}. \quad (4.7)$$

Dieser Kegel ist die Schnittmenge des Nullraumes der stöchiometrischen Matrix mit den positiven Halbräumen der irreversiblen Reaktionen. Ein Beispiel eines Flusskegels ist in Abbildung 4.1 dargestellt. Wie man der Abbildung entnehmen kann, sind die Kanten eines solchen Kegels von besonderer Bedeutung; sie werden daher in einem späteren Kapitel genauer betrachtet.

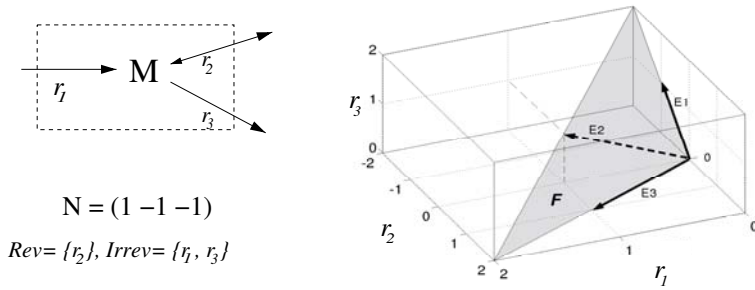


Abbildung 4.1: Beispiel eines Flusskegels. Der Kegel wird aufgespannt durch konvexe Kombinationen seiner Kanten (Extremalstrahlen E1 und E3) und ist in deren Richtung unbeschränkt. Die Kanten entsprechen dabei einer Teilmenge der Elementarmoden (E1, E2, E3) (siehe Kapitel 4.3.6).

Durch Hinzunahme der Nebenbedingungen C2 und C3 wird der Flusskegel auf eine kleinere Menge zulässiger Vektoren reduziert. Aus dem Flusskegel wird dann ein unbeschränktes oder beschränktes Polyeder (in letzterem Fall wird er auch als *Polytop* bezeichnet).

Im Folgenden werden nun zusätzlich verschiedene Optimalitätsbedingungen gefordert. Diese können grundsätzlich auch als Nebenbedingungen aufgefasst werden, da sie den zulässigen Lösungsraum noch weiter einschränken. Sie werden nachfolgend jedoch separat betrachtet, da sie das wesentliche Unterscheidungsmerkmal der verschiedenen Methoden darstellen.

4.3.3 Flussbilanzanalyse

Die Flussbilanzanalyse (FBA) ist eine der am häufigsten verwendeten constraint-basierten Methoden und beide Begriffe werden oft auch synonym verwendet. Einige spezielle Varianten der FBA wurden für spezifische Probleme formuliert und werden hier kurz vorgestellt. Das Ziel der Flussbilanzanalyse ist es, konkrete Flussverteilungen zu identifizieren, die sowohl die Nebenbedingungen der Klassen C1 bis C3 erfüllen als auch eine lineare Zielfunktion maximieren:

C4) Optimalität: $\mathbf{c}^T \mathbf{r} = c_1 r_1 + c_2 r_2 + \dots + c_q r_q = \text{Max!}$

Der Vektor $\mathbf{c}$ definiert das Optimalitätskriterium, indem die Beiträge der Reaktionsraten gewichtet werden. Zusammen mit den anderen Nebenbedingungen C1-C3 und C4 ergibt sich ein lineares Optimierungsproblem (Bertsimas und Tsitsiklis, 1997). Es sei darauf hingewiesen, dass zwar der Lösungswert der Zielfunktion eindeutig bestimmbar ist, es jedoch im Allgemeinen keine zugehörige eindeutige optimale Flussverteilung gibt, sondern vielmehr verschiedene Flussverteilungen dasselbe Optimum ergeben können.

Die am häufigsten verwendete *in silico* Zielfunktion ist die Maximierung der Biomasse-rate (bzw. des Wachstums). Für viele Mikroorganismen bzw. einzellige Lebewesen konnte gezeigt werden, dass unter bestimmten Umweltbedingungen (z.B. bei Substrat-Limitierung) das Prinzip der Biomassemaximierung als zelluläre Zielfunktion ein physiologisch realistisches Optimierungskriterium darstellt (Edwards et al., 2001; Feist et al., 2009; Ibarra et al., 2002). Eine alternative und biologisch plausible „natürliche“ Zielfunktion ist z.B. die Maximierung der ATP-Syntheserate. In (Schuetz et al., 2007) wird ein Überblick und eine Analyse weiterer Zielfunktionen gegeben. Eine „künstliche“ Zielfunktion, die insbesondere für biotechnologische Anwendungen von Bedeutung ist, ist die maximale Ausscheiderate eines Produktes (Varma et al., 1993). Es sei darauf hingewiesen, dass in vielen Fällen (durch die Festlegung der Substrataufnahmerate) effektiv die Produktausbeuten an Stelle der Produktraten optimiert werden. Die Unterscheidung zwischen Rate und Ausbeute ist für die FBA essentiell und wird in der Literatur intensiv diskutiert (Schuster et al., 2008; Teusink et al., 2009).

Für spezielle Anwendungen der FBA können einige (zum Teil zusätzlich eingeführte) Variablen in ihrem zulässigen Wertebereich auf ganzzahlige Werte beschränkt sein. So kann zum Beispiel die Verfügbarkeit einer bestimmten Reaktion mittels einer Hilfsvariablen, die auf die Werte 1 (für verfügbar) und 0 (für inaktiv) beschränkt ist, kodiert sein. Dies führt dann zu einem Problem der gemischt-ganzzahligen Optimierung (engl.: mixed integer linear programming, MILP), auf die später noch genauer eingegangen wird.

Varianten der Flussbilanzanalyse

Es wurde eine Reihe von speziellen Varianten der FBA entwickelt, die für bestimmte Aufgabenstellungen bzw. Anwendungen besonders geeignet sind. Dabei geht es meist um die Integration experimenteller Daten wie z.B. von Genexpressions-, Metabolitkonzentrations- oder Perturbationsdaten oder um die Beschreibung dynamischer (transienter) Änderungen der Flüsse. Ein umfassender Überblick über diese Methoden ist in (Lewis et al., 2012) gegeben. Ohne Anspruch auf Vollständigkeit sollen die wichtigsten Varianten im Folgenden kurz skizziert werden:

FBA mit Berücksichtigung der Regulation

Ein Ansatz, um (transkriptionelle) regulatorische Netzwerke mit FBA-Modellen zu kombinieren, wurde in (Covert et al., 2001) vorgeschlagen. Dabei wird ein Boole'sches Netzwerk mit genregulatorischen Bedingungen dem FBA-Modell vorgeschaltet. Das kombinierte rFBA-Modell kann den AN/AUS-Effekt bestimmter Umweltsignale auf die Expression der metabolischen Gene abbilden und somit auch die Verfügbarkeit bestimmter Pfade unter definierten Umweltbedingungen prädictieren. Obwohl rFBA nur Boole'sche Logik (z.B. Gen/Reaktion A ist aktiv, WENN Substrat S verfügbar ist UND Sauerstoff NICHT) und keine kausalen Zyklen berücksichtigen kann, wird die Genauigkeit der Prädiktion gegenüber reinen FBA-Modellen verbessert (Covert et al., 2004; Shlomi et al., 2007).

FBA mit Berücksichtigung von Metabolitkonzentrationen

Die Konsequenz der Fließgleichgewichtsannahme ist, dass Metabolitkonzentrationen und deren dynamischer Verlauf nicht berücksichtigt werden müssen. Genau diese Eigenschaft kann jedoch von Nachteil sein, wenn experimentelle Konzentrationsdaten vorliegen und integriert werden sollen. Da dies explizit nicht möglich ist, werden diese Daten oft mit Hilfe eines thermodynamischen Ansatzes integriert. Dabei wird die Bedingung genutzt, dass die freie Gibbs'sche Energie jeder Reaktion negativ sein muss, die in ihrer definierten Vorwärtsrichtung ablaufen soll. Durch die Integration der thermodynamischen Nebenbedingungen wird eine Reduzierung des Lösungsraumes erreicht (Hoppe et al., 2007). Dieser ist jedoch nicht mehr konvex, was die Suche nach gültigen Flussvektoren verkompliziert.

Eine weitere Möglichkeit, mit Hilfe thermodynamischer Nebenbedingungen den Lösungsraum einzuschränken, liegt in der Identifizierung und Beseitigung von thermodynamisch unzulässigen Zyklen. Unzulässige Zyklen sind stationäre Flussverteilungen ohne Verbrauch von externen Substraten oder Energie. Da die thermodynamischen Antriebskräfte sich in einem solchen Zyklus zu Null addieren, kann kein Netto-Fluss in einem solchen Zyklus auftreten. Diese Bedingung ist äquivalent zum zweiten Kirchhoff'schen Gesetz für elektrische Stromkreise. Diese Nebenbedingungen lassen sich ohne die explizite Berücksichtigung der Änderung der freien Gibbs'schen Energie als MILP-Probleme formulieren (Schellenberger et al., 2011).

FBA mit Berücksichtigung von Gen-Expressionsdaten

Es gibt zwar keinen einfachen Zusammenhang zwischen dem Expressionslevel eines enzymkodierenden Gens und dem zugehörigen Reaktionsfluss, jedoch sind einige Aussagen mit Hilfe simpler Heuristiken möglich. So kann in erster Näherung angenommen werden, dass niedrige Expressionslevel mit einem sehr kleinen Fluss korrelieren und hohe Expressionslevel dementsprechend als Indiz für große Flüsse gelten. Die oftmals zur Verfügung stehenden Expressionsdaten können so z.B. genutzt werden, um den Lösungsraum zu reduzieren und Gewebe- bzw. kontextspezifische Flüsse zu berechnen (iMAT-Methode (Shlomi et al., 2008)). Eine erweiterte Variante dieses Ansatzes wurde von (van Berlo et al., 2011) vorgeschlagen und Reviews ähnlicher Methoden für die Integration von Expressionsdaten in FBA-Studien sind in (Blazier und Papin, 2012; Reed, 2012) gegeben.

FBA und dynamische Flüsse

Hier ist das Ziel, die (dynamischen) Verläufe von Flüssen sowie von Konzentrationsprofilen ausgewählter Metabolite zu simulieren. Um den zeitlichen Verlauf der Konzentrationen von Biomasse und Metaboliten zu beschreiben, wurden verschiedene Methoden vorgeschlagen (Covert et al., 2001; Mahadevan et al., 2002; Varma und Palsson, 1994). Meist werden dabei zunächst mit FBA (und der steady-state Annahme für die internen Metabolite) die Austauschflüsse berechnet. Anschließend werden die Konzentrationsprofile der Biomasse und der externen Spezies (die explizit durch kinetische Gleichungen modelliert werden) durch Integration über diskrete Zeitschritte berechnet.

Ein weiterer Ansatz, iFBA (integrated FBA), bei dem Boolesche regulatorische Nebenbedingungen (rFBA) mit Differentialgleichungen für ausgewählte interne und externe Spezies gekoppelt werden, wurde von (Covert et al., 2008) vorgeschlagen. Mit diesem können Module bestehender Differentialgleichungs- und Boolescher Modelle von metabolischen und regulatorischen Prozessen mit FBA-Modellen integriert werden.

4.3.4 Flussvariabilitätsanalyse

Das Ziel der Flussvariabilitätsanalyse (FVA) ist es, unter gegebenen Nebenbedingungen die zulässigen Bereiche der Reaktionsraten zu bestimmen (Mahadevan und Schilling, 2003). Oftmals sind die exakten Werte nicht von entscheidender Bedeutung, so dass der Informationsgewinn aus berechneten Ober- und Untergrenzen der Flüsse für viele Fragestellungen ausreichend ist.

Die mathematische Formulierung ist der von FBA-Problemen sehr ähnlich. Zunächst seien wie zuvor Nebenbedingungen der Kategorien C1, C2 und C3 definiert. Wird durch den zulässigen Lösungsraum ein Polytop beschrieben (und nicht nur ein eindeutiger Punkt), so müssen einige Flüsse innerhalb gewisser Schranken variabel sein. Um diese Schranken zu identifizieren, werden für jede Reaktionsrate r_i zwei FBA-Probleme definiert. Dabei wird zuerst das Minimum und anschließend das Maximum bestimmt. Die entsprechenden Optimierungsprobleme sind demnach definiert durch:

$$\begin{aligned} \min_{\mathbf{r}} r_i \quad & \text{bzw.} \quad \max_{\mathbf{r}} r_i \\ \text{s.t.} \quad & \mathbf{Nr} = \mathbf{0} \\ & \alpha_i \leq r_i \leq \beta_i \\ & r_i = b_i \quad \text{für alle bekannten Raten } r_i. \end{aligned} \tag{4.8}$$

Da diese beiden FBA-Probleme für alle Reaktionen aufgestellt werden, sind insgesamt $2 \cdot$ (Anzahl Reaktionen) Optimierungsprobleme zu lösen. FVA kann somit ohne die Verwendung einer biologischen Zielfunktion benutzt werden, um den Raum zulässiger Flussverteilungen unter bestimmten Nebenbedingungen vollständig zu beschreiben. Wenn das Minimum und das Maximum einer Reaktionsrate zusammenfallen, so ist diese sogar eindeutig berechenbar. Durch die explizite Berücksichtigung der Nebenbedingungen aus Kategorie C2 können Reaktionsraten eindeutig berechenbar sein, die mittels MFA nicht bestimmbar wären. Die FVA ist eine einfache, aber sehr wertvolle Methode der constraint-basierten Modellierung. Sie ist daher ein oft benutztes Werkzeug für die Analyse unterbestimmter Netzwerke. Einige Anwendungsbeispiele können in (Reed und Palsson, 2004; Khannapho et al., 2008; Hädicke et al., 2011) gefunden werden.

4.3.5 Anwendungen der Flussbilanzanalyse

Im Folgenden werden einige Anwendungen der Flussbilanzanalyse im Hinblick auf die Identifizierung von Interventionen erläutert. Durch die Untersuchung struktureller Netzwerkeigenschaften können z.B. Hinweise auf geeignete Interventionen gewonnen werden. Des Weiteren liefert die Analyse der optimalen Flussverteilungen Erkenntnisse über maximale Produktausbeuten und kann genutzt werden, um z.B. den Bedarf an Reduktionsäquivalenten und ATP unter diesen Bedingungen zu bestimmen.

FBA zur Analyse funktionaler Netzwerkeigenschaften

Die FBA hat sich als ein sehr flexibler Modellierungsansatz etabliert, mit dem eine Reihe wichtiger funktionaler Eigenschaften von metabolischen Netzwerken analysiert werden können. Sie ist nicht durch die Größe der betrachteten Netzwerke limitiert, so dass auch genomskalige Netzwerke relativ einfach betrachtet werden können. So können zum Beispiel gekoppelte Reaktionen (Reaktionen, die immer gemeinsam aktiv bzw. inaktiv sein müssen) sowie essentielle oder blockierte Reaktionen Hinweise auf geeignete Überexpressions- oder Knockout-Kandidaten liefern (Burgard et al., 2004; David et al., 2011). Der Vorteil, diese Analysen mittels FBA und nicht mit Hilfe der Kernmatrix durchzuführen, ist darin zu sehen, dass die Reversibilitätsnebenbedingungen explizit berücksichtigt werden. Dadurch kann evtl. eine größere Menge an blockierten bzw. gekoppelten Reaktionen identifiziert werden als allein durch die Analyse der Kernmatrix. Zudem können so auch hierarchische Kopplungen aufgedeckt werden. Diese sind dadurch gekennzeichnet, dass eine Reaktion R1 immer dann aktiv ist, wenn auch eine Reaktion R2 aktiv ist; diese Beziehung jedoch nicht umgekehrt gilt, also R2 nicht immer aktiv sein muss, wenn R1 aktiv ist (Burgard et al., 2004; David et al., 2011). Mit Hilfe solcher Untersuchungen können implizite strukturelle Nebenbedingungen identifiziert werden, die auch Einschränkungen für die Regulation gekoppelter Reaktionen oder Pathways haben.

Prädiktion von essentiellen Reaktionen

Mittels der FBA-Methodik können Reaktionsdeletionen identifiziert werden, die zur Blockierung der Biomassesynthese führen. Dafür wird die entsprechende Reaktionsrate auf Null fixiert und das Maximum der Wachstumsrate bestimmt. Ist dieses Null, so handelt es sich um eine essentielle Reaktion. Auf diese Weise lassen sich auch für andere Netzwerkfunktionalitäten die essentiellen Reaktionen identifizieren.

In vielen Studien konnte gezeigt werden, dass die Vorhersagen der Lebensfähigkeit von Deletionsmutanten sehr gut mit experimentell beobachteten Phänotypen von Mikroorganismen korrelieren (z.B. (Edwards und Palsson, 2000a; Edwards et al., 2001; Becker und Palsson, 2008)). Auch aus Abweichungen zwischen Prädiktion und Beobachtung lassen sich wertvolle Informationen gewinnen. So impliziert eine falsch-positive Prädiktion (die Mutante ist nicht lebensfähig im Widerspruch zur Prädiktion) eine Falsifikation der Netzwerkstruktur, da offensichtlich zusätzliche alternative Pfade berücksichtigt werden müssen. Eine falsch-negative Prädiktion (die Mutante ist lebensfähig im Widerspruch zur Prädiktion) ist hingegen vermutlich auf eine unbekannte (unberücksichtigte) Einschränkung der Kapazität eines Pfades oder bestimmte regulatorische Nebenbedingungen zurückzuführen.

Prädiktion von optimalem Wachstumsverhalten

Wie bereits angeführt, zeigen einige Mikroorganismen wie z.B. *E. coli* ein stöchiometrisch optimales Verhalten in Bezug zur Biomasseausbeute. Dies gilt zumindest unter Substrat-limitierten Bedingungen (Edwards et al., 2001; Ibarra et al., 2002; Feist et al., 2009). Der zu erwartende Phänotyp kann mittels FBA und EM einfach bestimmt werden. Auch der Effekt genetischer Modifikationen auf das optimale Wachstumsverhalten kann einfach analysiert werden, indem Reaktions- bzw. Gendeletionen als zusätzliche Nebenbedingungen berücksichtigt werden. So kann z.B. untersucht werden, ob bestimmte Knockouts die maximale Biomasseausbeute reduzieren und andere Reaktionsraten einschränken (wie z.B. die Produktsynthese).

Prädiktion von Interventionseffekten

FBA kann benutzt werden, um die unmittelbaren Auswirkungen von Gen- bzw. Reaktionsdeletionen zu präzisieren. Zum Erreichen eines optimalen Wachstumszustandes nach einer Intervention bedarf es mehrerer Generationen unter dem Einfluss eines selektiven Drucks (siehe Kapitel 2.2). Die Flusslage direkt nach einer Perturbation zeigt meist kein wachstumsoptimales Verhalten. Eine Hypothese, wie die Zellen ihren Stoffwechsel schnell reorganisieren, ist die Annahme der „Minimization of Metabolic Adjustments“ (MoMA) (Segrè et al., 2002). Die biologische Argumentation hinter dieser Annahme ist, dass die Zellen versuchen, mit minalem Aufwand die neue Flussverteilung so ähnlich wie möglich der des Wildtypen anzunähern. Dieser Ansatz kann wie folgt formalisiert werden: Mit $\mathbf{w}$ als optimalem Flussvektor des Wildtypen und d als Index der deletierten Reaktionen ergibt sich :

$$\begin{aligned} \mathbf{N}\mathbf{r} &= \mathbf{0} \\ \alpha_i &\leq r_i \leq \beta_i \\ r_d &= 0 \\ (\mathbf{r} - \mathbf{w})^T (\mathbf{r} - \mathbf{w}) &= \text{Min!} \end{aligned} \tag{4.9}$$

Die ersten drei Terme entsprechen dabei den üblichen FBA Nebenbedingungen. Die Optimierung selbst ist in diesem Fall ein quadratisches Optimierungsproblem, das ebenfalls mit speziellen Solvern gelöst werden kann. Da die Lösungen von der Referenz-Flussverteilung abhängen, ist es zwingend erforderlich, dass diese bekannt ist. Sollte nur eine FBA-Lösung für den Wildtypen vorhanden sein, so besteht das Problem multipler optimaler Flussverteilungen für den Wildtypen und in der Folge auch für die berechneten Flussverteilungen der Mutanten. Für einige Mutanten von *E. coli* konnte gezeigt werden, dass dieser Ansatz zu verbesserten Prädiktionen (im Vergleich zur Standard-Zielfunktion) führt (Segrè et al., 2002). Ein anderer Algorithmus, dem eine ähnliche biologische Motivation zu Grunde liegt, ist ROOM (Regulatory On/Off Minimization) (Shlomi et al., 2005). Die Zielfunktion dieser Methode ist die Minimierung der *Anzahl* der Flüsse, die sich im Vergleich zum Wildtyp signifikant unterscheiden. Es müssen somit neben einer Referenzverteilung auch Grenzwerte definiert werden, die festlegen, ab welchem Ausmaß eine Flussänderung als signifikant beurteilt wird.

Prädiktion optimaler Produktausbeuten

Das Potenzial der Produktsynthese eines metabolischen Netzwerkes kann untersucht werden, indem für ein gegebenes Substrat die maximal erreichbaren Ausbeuten für alle synthetisierbaren Produkte des betrachteten Netzwerkes berechnet werden (evtl. unter verschiedenen Annahmen). Solche Analysen sind für biotechnologische Anwendungen und das Metabolic Engineering sehr nützlich (Stephanopoulos et al., 1998; Varma et al., 1993).

FVA zur Analyse optimaler Flussverteilungen

Eine Anwendung der FVA ist die Charakterisierung der gesamten optimalen Lösungsmenge eines FBA-Problems (Mahadevan und Schilling, 2003). Dafür wird der zuvor bestimmte optimale Wert w_{opt} der Zielfunktion $\mathbf{c}^T \mathbf{r}$ als zusätzliche Nebenbedingung zu den zuvor definierten Nebenbedingungen gesetzt ($\mathbf{c}^T \mathbf{r} = w_{opt}$). Anschließend wird die FVA-Methodik angewendet und für jede Reaktion das Intervall zulässiger Werte bestimmt. So kann z.B. durch Fixierung der Wachstumsrate auf das zuvor bestimmte Maximum der potentielle Bereich aller Flüsse unter wachstumsoptimalen Bedingungen identifiziert werden. In diesem Sinne wurden mittels FBA und FVA in (Hädicke et al., 2011) potentielle Eingriffsstellen identifiziert, die zu einer erhöhten Wasserstoffproduktion in schwefelfreien Purpurbakterien führen können.

4.3.6 Elementarmodenanalyse

Die Elementarmodenanalyse ist ein Werkzeug der metabolischen Pathway-Analyse, in der funktionale Teilnetze identifiziert und analysiert werden. Ein Pathway im biologischen Sinn ist dabei eine Reaktionsfolge, der eine bestimmte physiologische Funktion zugeordnet wird wie z.B. die Glykolyse oder der Zitronensäurezyklus. Es gibt verschiedene Ansätze für die mathematische Definition eines Pfades (z.B. (Leiser und Blum, 1987; Mavrovouniotis et al., 1990; Trinh et al., 2009)), wobei sich zwei Konzepte erfolgreich etabliert haben. Zum Einen sind dies die Extremen Pathways (Schilling et al., 2000) und zum Anderen die Elementarmoden (Schuster und Hilgetag, 1994; Schuster et al., 2000). Da diese beiden Ansätze sehr ähnlich sind (in vielen Fällen sogar identisch; für einen detaillierten Vergleich siehe (Klamt und Stelling, 2003; Papin et al., 2004)), werden im Folgenden nur die Elementarmoden weiter betrachtet.

Für die Definition der Elementarmoden (EM) wird zunächst das Konzept des Supports eines Vektors eingeführt. Der Support $supp(\mathbf{v})$ eines Vektors $\mathbf{v}$ ist definiert als die Menge der Reaktionsindizes, deren zugehörige Einträge für die Flusswerte ungleich Null sind: $supp(\mathbf{v}) = \{i | v_i \neq 0\}$.

Mit Hilfe dieser Definition lassen sich Elementarmoden als stationäre Flussvektoren beschreiben, die eine Support-minimale (nicht-reduzierbare) Menge von Reaktionen benutzen (Schuster und Hilgetag, 1994; Schuster et al., 2000) und die Bedingungen des Fließgleichgewichts erfüllen. Die exakte mathematische Definition ist dabei durch folgende drei Bedingungen gegeben (Schuster und Hilgetag, 1994; Schuster et al., 2000):

Ein Elementarmodus (EM) ist ein Flussvektor $\mathbf{e}$ mit folgenden Eigenschaften:

(i) Stationarität: $N\mathbf{e} = \mathbf{0}$

(ii) Reversibilitäts-Konformität: $e_i \geq 0 \quad (\forall i \in Irrev)$

(iii) Support-Minimalität (Elementarität, Unzerlegbarkeit): Es gibt keinen Vektor $\tilde{\mathbf{e}}$, der die Eigenschaften (i) und (ii) erfüllt und dessen Menge beteiligter Reaktionen eine Teilmenge von $\mathbf{e}$ darstellt $supp(\tilde{\mathbf{e}}) \subsetneq supp(\mathbf{e})$.

Ein EM $\mathbf{e}$ ist *reversibel* wenn $supp(\mathbf{e}) \cap Irrev = \emptyset$ und *irreversibel* anderenfalls.

Die Bedingungen (i) und (ii) sind identisch zu den Nebenbedingungen C1 und C2' aus den vorhergehenden Abschnitten zur FBA und FVA. Diese Bedingungen bilden den Flusskegel (Gleichung 4.7), in dem alle gültigen stationären Flussverteilungen liegen. Bedingung (iii), die auch als *genetische Unabhängigkeit* bezeichnet wird, garantiert, dass ein EM eine minimale Anzahl von Reaktionen verwendet, so dass keine Teilmenge von aktiven Reaktionen eines EM eine nicht triviale ausgeglichene Flussverteilung generieren kann.

Mit den Bedingungen (i)-(iii) ist die Menge der EM eines Netzwerkes eindeutig definiert (bis auf einen Skalierungsfaktor für jeden EM). Zudem können alle zulässigen stationären Flussvektoren $\mathbf{r}$ durch die EM generiert werden (Gleichung (4.10)). Die Zerlegung eines konkreten Flussvektors $\mathbf{r}$ in EM ist im Allgemeinen jedoch nicht eindeutig. Für einige Anwendungen von EM wäre eine eindeutige Zerlegung dennoch wünschenswert, so dass einige Heuristiken zu diesem Zweck vorgeschlagen wurden (Ip et al., 2011; Wiback et al., 2003).

Abb. 4.2 zeigt die drei Elementarmoden des Beispielnetzwerkes N_1 und die zugehörigen relativen Flüsse (die Rate der inaktiven Reaktionen ist Null). Die EM wurden auf eine Substrataufnahmerate von 1 normiert. Es ist gut ersichtlich, dass die EM minimale funktionale Teilnetzwerke repräsentieren. In N_1 wandeln die EM das externe Substrat S in die Produkte P1, P2 und P3 um. Die internen Metabolite A bis D werden dabei weder akkumuliert noch verbraucht, da für jeden Metaboliten gilt, dass die Summe der produzierenden und verbrauchenden Flüsse gleich ist.

EM werden meist mit Hilfe der *double description method* berechnet, einem bekannten Algorithmus aus dem Gebiet der algorithmischen Geometrie (Fukuda und Prodon, 1996). Dabei wird ein Tableau verwendet, bei dem durch iterative Anwendung von Gaußschen Kombinationen neue Kandidatenvektoren generiert werden. Der aufwändige Teil des Algorithmus ist die Überprüfung dieser Kandidaten auf Elementarität. Die Identifizierung aller EM ist ein kombinatorisches Problem und kann sich als schwierig darstellen, da in größeren Netzwerken (> 100 Reaktionen) schnell mehrere Millionen EM berechnet werden müssen. Durch einige algorithmische Verbesserungen in jüngster Zeit (Gagneur und Klamt, 2004; Klamt et al., 2005; Terzer und Stelling, 2008; Urbanczik und Wagner, 2005) ist es nun durchaus möglich, mehrere hundert Millionen EM zu berechnen. Dennoch kann die vollständige Menge aller EM in genomskaligen Netzwerken häufig nicht

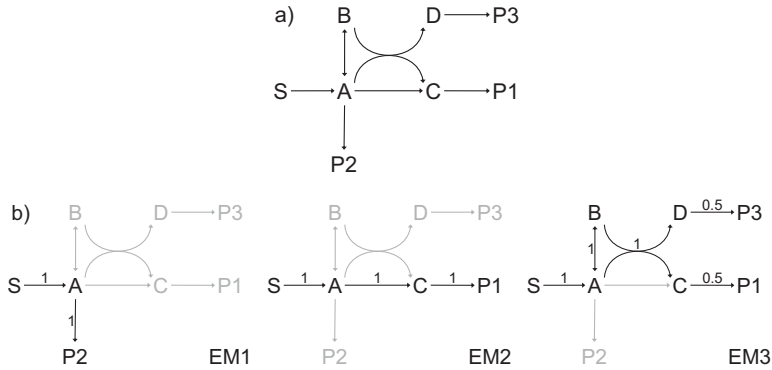


Abbildung 4.2: a) Beispielnetzwerk N_1 . b) Elementarmoden und zugehörige relative Flüsse. Hellgrau hinterlegt: inaktive Reaktionen

berechnet werden. In solchen Fällen können trotzdem zumindest die kürzesten EM bestimmt werden (de Figueiredo et al., 2009) oder Projektionsmethoden angewandt werden (Marashi et al., 2012). Ein alternativer Ansatz, das Sampling von Teilmengen von EM, wird in einem späteren Kapitel vorgestellt.

Elementarmoden besitzen einige theoretische Eigenschaften, die sich als sehr nützlich für die metabolische Netzwerkanalyse herausgestellt haben. Die folgenden drei Eigenschaften sind dabei ganz besonders wertvoll:

Darstellung von EM als Reaktionsmengen

Ein EM wird meist als Vektor $\mathbf{e}$ repräsentiert (siehe Definition oben). Da ein EM jedoch auch eindeutig durch seinen Support definiert ist, ist es ebenso möglich, ihn als Menge seiner beteiligten Reaktionen darzustellen $E = \text{supp}(\mathbf{e})$. Diese Darstellung ist besonders vorteilhaft, wenn mit kombinatorischen Eigenschaften von EM-Mengen gearbeitet wird. Die stöchiometrischen Koeffizienten eines EM können, falls sie in einer bestimmten Anwendung benötigt werden, einfach aus der Mengen-Repräsentation und der stöchiometrischen Matrix berechnet werden (Gagneur und Klamt, 2004).

Verbleibende EM nach Entfernen von Reaktionen

Wenn eine Reaktion in einem Netzwerk gelöscht wird, so ist die Menge der EM in dem verbleibenden Teilnetzwerk (das nur eine Teilmenge S der Reaktionen des gesamten Netzwerkes enthält) gegeben durch die Teilmenge aller EM des Originalnetzwerkes, in denen die gelöschte Reaktion nicht auftritt. Diese können ohne Neuberechnung einfach selektiert werden, indem alle EM ausgewählt werden, deren Support vollständig in S enthalten ist: $\text{supp}(\mathbf{e}) \subset S$.

EM generieren den Flusskegel

Linearkombinationen von EM (mit nicht-negativen Koeffizienten für irreversible EM und beliebigen Koeffizienten für reversible EM) bilden den Flusskegel $\mathcal{F}$ und bieten damit eine alternative Beschreibung für Gleichung (4.7):

$$\mathcal{F} = \{\mathbf{r} \in \mathbb{R}^q \mid \mathbf{r} = \sum \alpha_j \mathbf{e}^j, \alpha_j \geq 0 \text{ if } \mathbf{e}^j \text{ irreversible}\}. \quad (4.10)$$

Die letzte Eigenschaft verdeutlicht, dass die unendlich vielen Lösungen des Nullraumes der stöchiometrischen Matrix auf eine endliche Menge von zulässigen, qualitativ verschiedenen Lösungen abgebildet wird. Die Bestimmung dieser verschiedenen optimalen und suboptimalen Stoffwechseloptionen (z.B. bzgl. eines Ausbeutekriteriums) ermöglicht eine systematische Untersuchung der metabolischen Phänotypen. So können zum Beispiel aus Eigenschaften, die allen EM mit einem gewünschten Verhalten gemeinsam sind, geeignete Interventions-Kandidaten bestimmt werden. Darüber hinaus können Netzwerkeigenschaften wie z.B. die Struktur, Regulierung, Robustheit und Fragilität analysiert werden (Trinh et al., 2009). Des Weiteren ergeben sich Anwendungen für ein zielgerichtetes Stammdesign (siehe Kapitel 4.4.2 und 5.1).

4.3.7 Minimal Cut Sets

Wie bereits oben erläutert, beschreibt das (Un-)Gleichungssystem aus C1 und C2' die Menge der zulässigen stationären Flussverteilungen. Die zugehörigen EM sind gültige Lösungen dieses Systems und generieren den Flusskegel. Die oben erwähnten Eigenschaften der EM erlauben eine einfache Analyse der Auswirkungen von Reaktionsdeletionen. Im Folgenden wird für einen einzelnen EM $\mathbf{e}$ die Repräsentation mittels seines Supports verwendet $E = \text{supp}(\mathbf{e})$. Die vollständige Menge der EM wird durch $\mathbf{E}$ beschrieben.

Wenn eine bestimmte Flussverteilung oder Netzwerkfunktion durch das Entfernen von dafür geeigneten Reaktionen (bzw. den entsprechenden Genen) ausgeschlossen werden sollen, ist es nützlich, die unerwünschten Flussverteilungen mit der Menge der zugehörigen EM $\mathbf{T}$ zu beschreiben ($\mathbf{T} \subset \mathbf{E}$). Diese Menge $\mathbf{T}$ von EM wird in diesem Zusammenhang als Target- oder Zielmoden bezeichnet, wobei das „Ziel“ hier die unerwünschte Funktion ist, die es zu entfernen gilt. Ein Cut Set C ist definiert als eine Reaktionsmenge, so dass jeder EM aus $\mathbf{T}$ mindestens eine dieser Reaktionen enthält:

$$\forall T \in \mathbf{T} : C \cap T \neq \emptyset \quad (4.11)$$

Daraus folgt, dass das Entfernen der Reaktionen eines Cut Sets die Aktivität aller Zielmoden unterbindet, da per Definition, durch das Entfernen von mindestens einer Reaktion in jedem Target-EM letztere keine stationäre Flussverteilung mehr etablieren können. Analog zur Elementaritätsbedingung für die EM kann auch für die Cut Sets eine Minimalitätsforderung gestellt werden (Klamt und Gilles, 2004): Ein Cut Set C ist ein minimaler Cut Set (MCS), wenn es keine Menge $\tilde{C}$ gibt, die eine Teilmenge von C darstellt ($\tilde{C} \subsetneq C$) und Gleichung 4.11 erfüllt. Dieser Definition folgend können die MCS auch als „minimal hitting sets“ der Targetmoden interpretiert werden (Klamt, 2006), die wiederum mit dem Berge-Algorithmus (Berge, 1989) berechnet werden können.

Es sei für das zuvor betrachtete Beispielnetzwerk N_1 das Ziel, ausschließlich P1 mit optimaler Ausbeute zu produzieren (Abb. 4.2). Dies kann nur durch EM2 gewährleistet werden, so dass die Elementarmoden EM1 und EM3 eliminiert werden müssen. Die Menge der Target-Moden ist demnach durch $T = \{EM1, EM3\}$ gegeben. Es existieren fünf MCS mit denen beide EM deletiert werden können; diese sind in Abb. 4.3 dargestellt. Jeder dieser MCS „trifft“ sowohl EM1 als auch EM3, jedoch erfüllt jeweils keine Teilmenge eines der MCS diese Eigenschaft.

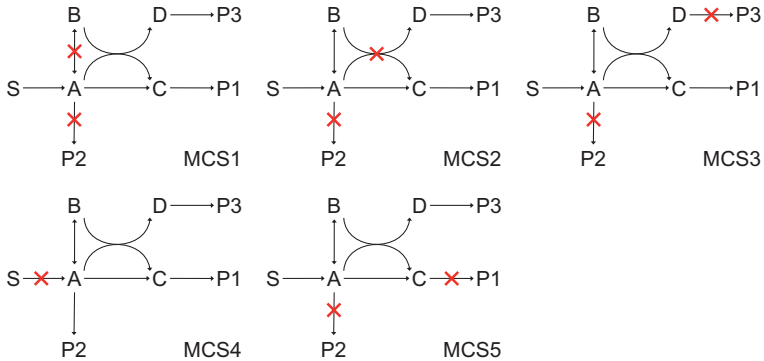


Abbildung 4.3: Minimal Cut Sets für die Unterdrückung der suboptimalen Produktion von P1 im Beispielnetzwerk N_1 . Die Target-EM entsprechen den Elementarmoden EM1 und EM3 aus Abb. 4.2

Im Prinzip können viele MCS-Probleme auch ohne explizite Berechnung der EM und Definition von Targetmoden durch eine geeignete Formulierung als FBA-Problem gelöst werden. So haben (Suthers et al., 2009) für ihren Ansatz der „Synthetischen Lethalität“ durch systematische Simulation aller Einer-, Zweier- und Dreierkombinationen von Reaktionsdeletionen diejenigen MCS bestimmt, die zur Unterdrückung der Biomasse-Produktion in *E. coli* führen. Sofern die vollständige Menge der EM berechnet werden kann, hat sich der Ansatz, zuerst die EM und anschließend die minimal hitting sets zu berechnen, in der Praxis jedoch als effizienter erwiesen (Haus et al., 2008). Zudem ist es für komplexe Mengen von Zielmoden manchmal schwierig bis unmöglich, ein entsprechendes FBA-Problem zu formulieren.

Anwendungen der MCS umfassen z.B. Analysen der strukturellen Fragilität bzw. Robustheit, wobei durch die MCS kleinste Fehlermengen (z.B. inaktivierte Reaktionen, die zu unerwünschten Netzwerkeigenschaften führen) abgebildet und untersucht werden können (Behre et al., 2008). Des Weiteren können gezielte Reaktionsdeletionen bestimmt werden, mit denen unerwünschte Netzwerkeigenschaften unterdrückt werden können (Klamt und Gilles, 2004; Klamt, 2006).

Dualität von EM und MCS

Elementarmoden und Minimal Cut Sets sind zwei duale Konzepte der Netzwerkanalyse (Ballerstein et al., 2012). So lassen sich nicht nur die MCS als „hitting sets“ der EM berechnen, sondern es können auch die EM als „hitting sets“ der MCS berechnet werden. Dafür müssten jedoch zuvor die MCS unabhängig von den EM berechnet werden. Theoretisch ist es durchaus möglich, die MCS ohne Kenntnis der zugehörigen EM zu berechnen, indem durch Exhaustionsmethoden systematisch alle möglichen Reaktion-Deletions-Kombinationen mit dem FBA-Ansatz auf ihre Funktionalität hin untersucht werden. Eine elegantere, alternative Lösung ist durch die Formulierung geeigneter MILP-Probleme gegeben (Suthers et al., 2009). Diese direkten Ansätze stoßen rechentechnisch jedoch schnell an ihre Grenzen, so dass schon für Netzwerke mittlerer Größe nicht mehr alle MCS in vertretbarer Zeit enumeriert werden können. Dennoch wird durch die unabhängige Berechnung belegt, dass EM und MCS äquivalente Beschreibungen der Netzwerkfunktionalität sind, wenn auch ausgehend von unterschiedlichen Perspektiven.

Die Dualität zwischen EM und MCS geht weit über die Möglichkeit der gegenseitigen Berechnung hinaus. So wurde kürzlich gezeigt, dass für ein Netzwerk mit gegebener Funktionalität ein duales Netzwerk konstruiert werden kann, in dem die EM (MCS) den MCS (EM) des originalen Netzwerkes entsprechen (Ballerstein et al., 2012).

Diese Dualitätseigenschaften eröffnen neue algorithmische Möglichkeiten, EM oder MCS zu berechnen (von Kamp und Klamt, 2014; Ballerstein et al., 2012). Durch Ausnutzung der Dualitätseigenschaft konnten zum Beispiel in (von Kamp und Klamt, 2014) alle MCS (die das Wachstum unterdrücken) bis zur Kardinalität fünf für ein genomskaliges Netzwerk berechnet werden. Dafür wurde zunächst das zugehörige duale Netzwerk konstruiert und anschließend ein Algorithmus zur Berechnung der kürzesten Elementarmoden verwendet. Auf diese Art können tausende MCS enumeriert werden, aus denen in einem späteren Schritt die geeignetsten Kandidaten für mögliche Interventionen selektiert werden können. Abbildung 4.4 fasst die Dualitätsprinzipien (bio)chemischer Netzwerke zusammen.

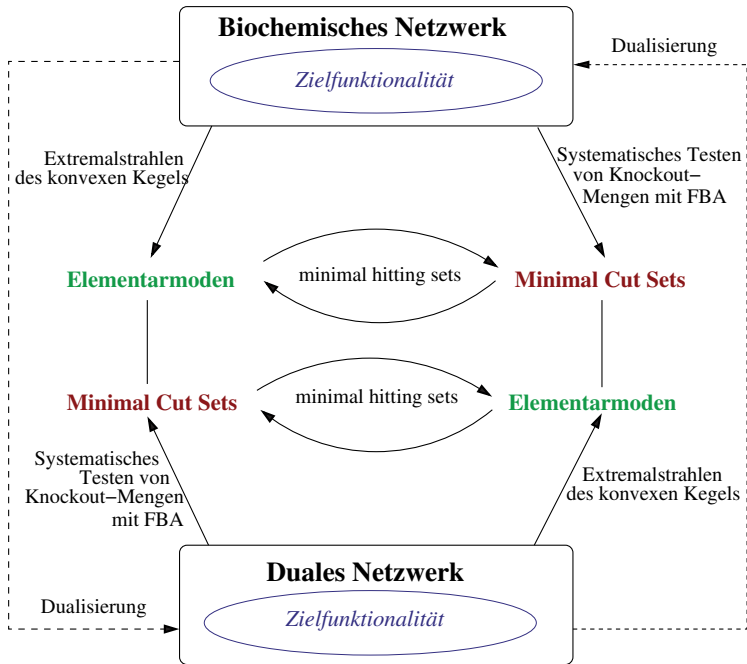


Abbildung 4.4: Dualitätsprinzipien von EM und MCS.

4.4 Modellbasierte Methoden zur Identifizierung von Interventionsstrategien

In der letzten Dekade sind verschiedene mathematische Ansätze zur modellbasierten Identifikation von Interventionsstrategien zum Metabolic Engineering entwickelt worden. Durch geeignete Anwendungen constraint-basierter Methoden wie MFA, FBA oder EM-Analysen können bereits nützliche Analysen wie z.B. die Charakterisierung optimaler Flussverteilungen oder die Bestimmung maximaler Produktausbeuten im Hinblick auf das Stammdesign vorgenommen werden. Des Weiteren wurde eine Klasse spezieller Methoden entwickelt, deren Ziel die Identifizierung konkreter (genetischer) Interventionsstrategien ist. Im folgenden Abschnitt wird ein Überblick über diese Methoden gegeben, wobei der Fokus auf den stöchiometrischen Ansätzen liegt.

4.4.1 Ansätze mittels mathematischer Optimierung

Die im Folgenden diskutierten Methoden unterscheiden sich hauptsächlich in der verwendeten Zielfunktion und den integrierten Nebenbedingungen. Allen gemeinsam ist das Ziel, unter wachstumsoptimalen Bedingungen (gemeint ist hier meist die Optimalität bzgl. der Biomasseausbeute) eine Zielfunktion bzgl. der Produktsynthese zu optimieren.

OptKnock

Der erste constraint-basierte Algorithmus mit dem expliziten Ziel, Knockout-Strategien zu identifizieren, die zu einer gesteigerten Produktsynthese führen, wurde 2003 von Burgard et al. vorgestellt (Burgard et al., 2003). Der Idee von OptKnock liegt die Beobachtung zu Grunde, dass Mikroorganismen auf externe Änderungen reagieren, indem sie ihre Stoffflüsse so umlenken, dass auch unter neuen Bedingungen das realisierbare optimale Wachstumsverhalten erreicht wird (in Bezug auf die Biomasseausbeute bzw. Wachstumsrate) (Ibarra et al., 2002). Somit muss jede Intervention so ausgelegt werden, dass deren Auswirkungen unter dem neuen optimalen Wachstumsverhalten zum gewünschten Ergebnis führen. Das Netzwerk sollte demnach so umgestaltet werden, dass der gewünschte Stoff als obligatorisches Nebenprodukt der Biomassesynthese ausgeschieden wird, wenn der Organismus optimales Wachstumsverhalten erreicht (siehe Kapitel 2.2).

Zu diesem Zweck wird ein zweistufiges MILP-basiertes Optimierungsproblem formuliert. In einer äußeren Zielfunktion wird die gewünschte „technische“ Zielfunktion, üblicherweise die Ausscheidereaktion des Produktes, mittels Gen- oder Reaktions-Knockouts optimiert. Die maximale Anzahl der Knockouts wird dabei durch eine Nebenbedingung vorgegeben. In einer inneren Zielfunktion wird das biologische Verhalten der Zelle simuliert. Meist wird für diese Zielfunktion die Maximierung der Wachstumsrate (bzw. Biomasseausbeute) angenommen. Bei der Maximierung der inneren Zielfunktion müssen zusätzlich alle üblichen Nebenbedingungen wie z.B. die Stationarität, Irreversibilitätsbedingungen und Grenzen für bestimmte Flüsse (z.B. maximale Substrataufnahmeraten) erfüllt werden. Die mathematische Beschreibung ist damit durch folgendes MILP-Problem gegeben:

Maximiere y_i	$r_{\text{technisch}}$										
subject to	<table> <tr> <td>Maximiere r_i</td><td>r_{Biomasse}</td></tr> <tr> <td>subject to</td><td> Nr = 0 $r_{\text{Biomasse}} \geq r_{\text{Biomasse}}^{\text{target}}$ // Mindest-Biomasse $r_{\text{Aufnahme}} = r_{\text{Aufnahme}}^{\text{target}}$ // Substrat-Aufnahme $r_i^{\text{min}} \leq r_i \leq r_i^{\text{max}} \forall i \notin M$ // Weitere Schranken $r_i^{\text{min}} \cdot y_i \leq r_i \leq r_i^{\text{max}} \cdot y_i$ // Knockouts </td></tr> <tr> <td>$y_i = \{0, 1\}$</td><td>$\forall i \in M$ // zulässige Kandidaten</td></tr> <tr> <td>$\sum_{i \in M} (1 - y_i) \leq K$</td><td>// Beschränkung der</td></tr> <tr> <td></td><td>// Anzahl der Knockouts</td></tr> </table>	Maximiere r_i	r_{Biomasse}	subject to	Nr = 0 $r_{\text{Biomasse}} \geq r_{\text{Biomasse}}^{\text{target}}$ // Mindest-Biomasse $r_{\text{Aufnahme}} = r_{\text{Aufnahme}}^{\text{target}}$ // Substrat-Aufnahme $r_i^{\text{min}} \leq r_i \leq r_i^{\text{max}} \forall i \notin M$ // Weitere Schranken $r_i^{\text{min}} \cdot y_i \leq r_i \leq r_i^{\text{max}} \cdot y_i$ // Knockouts	$y_i = \{0, 1\}$	$\forall i \in M$ // zulässige Kandidaten	$\sum_{i \in M} (1 - y_i) \leq K$	// Beschränkung der		// Anzahl der Knockouts
Maximiere r_i	r_{Biomasse}										
subject to	Nr = 0 $r_{\text{Biomasse}} \geq r_{\text{Biomasse}}^{\text{target}}$ // Mindest-Biomasse $r_{\text{Aufnahme}} = r_{\text{Aufnahme}}^{\text{target}}$ // Substrat-Aufnahme $r_i^{\text{min}} \leq r_i \leq r_i^{\text{max}} \forall i \notin M$ // Weitere Schranken $r_i^{\text{min}} \cdot y_i \leq r_i \leq r_i^{\text{max}} \cdot y_i$ // Knockouts										
$y_i = \{0, 1\}$	$\forall i \in M$ // zulässige Kandidaten										
$\sum_{i \in M} (1 - y_i) \leq K$	// Beschränkung der										
	// Anzahl der Knockouts										

Eine Limitierung des OptKnock-Ansatzes ist darin zu sehen, dass bei wachstumsoptimalem Verhalten nur das Maximum der Produktausscheidung betrachtet wird (es wird implizit der optimistischste Fall angenommen). So kann es unter Umständen passieren, dass das Minimum der Produktausscheiderate unter diesen Bedingungen sehr kleine Werte (im Extremfall genau 0) annehmen kann. Da *in vivo* sowohl das Maximum als auch das Minimum von der Zelle realisiert werden kann, ist das von OptKnock prognostizierte Maximum der Produktausbeute gerade nicht obligatorisch an die Biomassessynthese gekoppelt (siehe Abb. 4.5). Dennoch bereitete OptKnock die Basis für viele weitere Methoden, die alle auf dessen zu Grunde liegender Idee aufbauen.

RobustKnock

RobustKnock (Tepper und Shlomi, 2010) ist eine Modifizierung des OptKnock-Ansatzes, bei der die Knockout-Kandidaten so ausgewählt werden, dass die mögliche Entkopplung der Produkt- und Biomassessynthese vermieden wird. Dafür wird das Minimum der Produktausscheidung bei wachstumsoptimalem Verhalten maximiert, so dass hier im Gegensatz zur OptKnock-Formulierung der „worst case“ angenommen wird. Die identifizierten Knockouts garantieren damit bei wachstumsoptimalem Verhalten eine bestimmte Mindestproduktion des gewünschten Stoffes. Mathematisch wird dies über ein dreistufiges MILP-Optimierungsproblem gelöst, wobei das äußere max-min Problem das Minimum der Produktausscheiderate maximiert und das innere Problem identisch zu dem von OptKnock ist. Auf die vollständige mathematische Formulierung wird hier verzichtet, der Aufbau des Optimierungsproblems kann wie folgt skizziert werden:

Maximiere durch Knockouts	<table> <tr> <td>Minimiere durch Flüsse</td><td>Technische Zielfunktion</td></tr> <tr> <td>subject to</td><td> <table> <tr> <td>Maximiere durch Flüsse</td><td>Biologische Zielfunktion</td></tr> <tr> <td>subject to</td><td>Steady-State</td></tr> <tr> <td></td><td>Mindestbiomasse</td></tr> <tr> <td></td><td>Substrataufnahme</td></tr> <tr> <td></td><td>Knockouts</td></tr> </table> </td></tr> </table>	Minimiere durch Flüsse	Technische Zielfunktion	subject to	<table> <tr> <td>Maximiere durch Flüsse</td><td>Biologische Zielfunktion</td></tr> <tr> <td>subject to</td><td>Steady-State</td></tr> <tr> <td></td><td>Mindestbiomasse</td></tr> <tr> <td></td><td>Substrataufnahme</td></tr> <tr> <td></td><td>Knockouts</td></tr> </table>	Maximiere durch Flüsse	Biologische Zielfunktion	subject to	Steady-State		Mindestbiomasse		Substrataufnahme		Knockouts
Minimiere durch Flüsse	Technische Zielfunktion														
subject to	<table> <tr> <td>Maximiere durch Flüsse</td><td>Biologische Zielfunktion</td></tr> <tr> <td>subject to</td><td>Steady-State</td></tr> <tr> <td></td><td>Mindestbiomasse</td></tr> <tr> <td></td><td>Substrataufnahme</td></tr> <tr> <td></td><td>Knockouts</td></tr> </table>	Maximiere durch Flüsse	Biologische Zielfunktion	subject to	Steady-State		Mindestbiomasse		Substrataufnahme		Knockouts				
Maximiere durch Flüsse	Biologische Zielfunktion														
subject to	Steady-State														
	Mindestbiomasse														
	Substrataufnahme														
	Knockouts														
subject to Anzahl Knockouts $\leq$ Limit															

Die verschiedenen Auswirkungen des OptKnock- und RobustKnock-Konzeptes können anhand der resultierenden Phänotypebenen diskutiert werden (Abbildung 4.5). Die durch die gestrichelte blaue Linie begrenzte Phänotypebene ist typisch für eine Mutante die mittels OptKnock konstruiert wurde. Wenn diese Mutante wachstumsoptimales Verhalten zeigt, kann die Produktausbeute jeden Wert zwischen den beiden blauen Punkten annehmen. Der obere Punkt beschreibt dabei die potentielle optimale Produktausbeute $(Y_{max}^{P/S}|\mu_{max})_{OptKnock}$ und der untere die ebenfalls zulässige minimale Produktausbeute $(Y_{min}^{P/S}|\mu_{max})_{OptKnock}$ unter wachstumsoptimalen Bedingungen.

Der Bereich, der durch die gestrichelte rote Linie begrenzt ist, stellt eine exemplarische Phänotypebene einer mittels RobustKnock identifizierten Mutante dar. Die optimale Produktausbeute $(Y_{max}^{P/S}|\mu_{max})_{RobustKnock}$ unter wachstumsoptimalen Bedingungen (roter Punkt) kann kleiner als die optimale Ausbeute der OptKnock-Mutante sein. Allerdings ist die minimale Produktausbeute bei optimalem Wachstum größer als die minimale Produktausbeute der OptKnock-Mutante unter diesen Bedingungen. Entscheidend ist, dass RobustKnock höhere Produktausbeuten unter wachstumsoptimalen Bedingungen garantieren kann:

$$(Y_{min}^{P/S}|\mu_{max})_{OptKnock} \leq (Y_{min}^{P/S}|\mu_{max})_{RobustKnock} \leq (Y_{max}^{P/S}|\mu_{max})_{OptKnock}$$

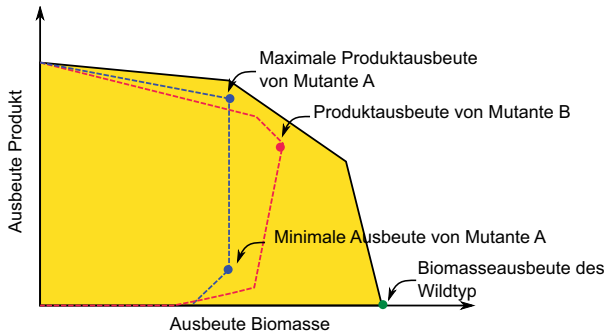


Abbildung 4.5: Hypothetische Phänotypebene eines Wildtyporganismus (gelber Bereich) und von zwei möglichen Knockout-Mutanten. Blaue gestrichelte Linie: Phänotyp-Ebene einer Mutante mit Realisierung einer OptKnock-Strategie; rote gestrichelte Linie: Phänotypebene einer Mutante mit Realisierung einer RobustKnock-Strategie.

Ein Nachteil beider Ansätze ist, dass die erreichte Kopplung sehr sensitiv bezüglich der Annahme der optimalen Biomasseausbeute sein kann. Falls die Mutanten sich nicht wachstumsoptimal verhalten, kann z.B. keine der beiden Mutanten in Abbildung 4.5 eine Mindestausbeute für das Produkt garantieren, so dass unter Umständen kein Produkt ausgeschieden wird.

OptGene

Werden die beiden zuvor beschriebenen Methoden in genomskaligen Netzwerken angewendet, sind durch den sehr großen Suchraum und den rechentechnischen Aufwand die identifizierten Strategien meist auf drei bis vier Knockouts beschränkt. Zudem wird jeweils nur genau eine optimale Strategie identifiziert und die Bestimmung weiterer suboptimaler Lösungen ist rechentechnisch sehr aufwändig (zuvor bestimmte Lösungen müssen mittels Nebenbedingungen ausgeschlossen werden). Diese Probleme versucht OptGene (Patil et al., 2005) zu umgehen, indem ein genetischer Optimierungsalgorithmus zur Identifizierung geeigneter Knockouts verwendet wird.

Da die genetische Optimierung unabhängig von den gewählten Zielfunktionen ist, können auch nichtlineare äußere (wie z.B. die gekoppelte Ausbeute von Produkt und Biomasse (BPCY), definiert als $BPCY = \mu \cdot r_{Produkt}$) und innere (wie z.B. MOMA (Segrè et al., 2002) oder ROOM (Shlomi et al., 2005)) Zielfunktionen berücksichtigt werden. OptGene kann nicht garantieren, das globale Optimum des Optimierungsproblems zu identifizieren. Dennoch können zumindest mehrere lokale, oft suboptimale Lösungen, die eine deutliche Verbesserung im Vergleich zum Wildtyp darstellen, gefunden werden. Das Vorgehen der Methode kann wie folgt zusammengefasst werden:

-
1. Wahl einer Startpopulation von Individuen, wobei ein Individuum charakterisiert wird durch die Menge seiner aktiven und inaktiven Reaktionen.
 2. Mit Hilfe der biologischen (inneren) Zielfunktion wird die Flussverteilung jedes Individuums berechnet.
 3. Jedem Individuum wird anschließend durch Auswerten der technischen Zielfunktion ein Fitness-Score zugeteilt.
 4. Wenn die Fitness-Scores ein bestimmtes Kriterium erfüllen → STOP.
Wenn nicht, dann werden unter Berücksichtigung bestimmter Regeln (Rekombinationen, Mutationen) neue Individuen generiert und mit 2. fortgesetzt.
-

OptReg

OptReg (Pharkya und Maranas, 2006) ist eine Erweiterung der OptKnock-Idee, die ebenfalls auf der Formulierung als zweistufiges Optimierungsproblem mit innerer und äußerer Zielfunktion basiert. Zusätzlich wird hier in der inneren Zielfunktion verlangt, dass der absolute Gesamtfluss möglichst gering wird. Zudem werden neben Knockouts auch Hoch- und Runterregulierungen von Reaktionen erlaubt, so dass drei verschiedene Mengen von binären Variablen definiert werden müssen. Eine Reaktion wird dabei als überexprimiert bzw. reprimiert betrachtet, wenn der Fluss durch die Reaktion der Mutante deutlich größer bzw. kleiner als beim Wildtyp ist. In dieser Klassifikation ist allerdings auch die Limitierung dieser Methode zu sehen, da die Ergebnisse von der verwendeten Referenzflusslage abhängen und diese, insbesondere in großen Netzwerken, nur selten gegeben ist.

OptStrain und SimOptStrain

OptStrain (Pharkya et al., 2004) war die erste Methode, mit der zusätzlich zu Knockout-Kandidaten heterologe Reaktionen bzw. Gene identifiziert werden konnten, die in dem betrachteten Stamm zu verbesserter Produktausbeute führen. In einem zweistufigen Vorgehen wird zunächst die optimale Ausbeute unabhängig vom Produktionsstamm mit Hilfe einer universellen Datenbank (mit mehr als 5700 Reaktionen) bestimmt (mit einem gegebenen Substrat). Das Netzwerk des Produktionsstammes wird systematisch mit Reaktionen aus dieser Datenbank so erweitert, dass die optimale Ausbeute mit einer minimalen Anzahl heterologer Reaktionen erreicht wird. Anschließend wird im zweiten Schritt die OptKnock-Routine angewendet, um die heterologen Pfade obligatorisch an den Stoffwechsel des Produktionsstammes zu koppeln, wodurch die realisierbaren (gekoppelten) Ausbeuten kleiner als die des Optimums des universellen Netzes sind.

SimOptStrain (Kim et al., 2011) ist eine Variante bzw. Erweiterung dieses Ansatzes, bei dem die Knockout-Kandidaten und heterologen Reaktionen (aus der selben Datenbank) simultan identifiziert werden. Die identifizierten Strategien können zu höheren gekoppelten Ausbeuten führen (im Vergleich zu OptStrain Strategien), obwohl die ungekoppelten Ausbeuten kleiner als die des Optimums des universellen Netzes sein können. Eine vereinfachte schematische Beschreibung beider Verfahren ist gegeben durch:

OptStrain		SimOptStrain	
1. Addition der minimalen Anzahl an Reaktionen, so dass die theoretische Produktausbeute maximiert wird.		1. max KO, RA s.t.	
2. max KO s.t.		r_{Engineering} max Flüsse s.t. r_{Biomasse} Steady-State Mindestbiomasse Substrataufnahme weitere Schranken Knockouts Neue Reaktionen	
max Flüsse s.t. Steady-State Mindestbiomasse Substrataufnahme weitere Schranken Knockouts		#KO ≤ K #KO ≤ K #RA ≤ L GPRs	

KO: Knockouts, RA: Reaktionsadditionen, GPRs: Gen-Protein-Reaktions-Assoziationen,

#: Anzahl

OptOrf

OptORF (Kim und Reed, 2010) basiert auf der grundlegenden Idee von OptKnock und ist als zweistufiges MILP-Problem formuliert. Im Gegensatz zu den bisherigen Methoden werden Gen-Protein-Reaktions-Assoziationen (GPRs) explizit berücksichtigt, so dass die Interventionen nicht für Reaktionen, sondern direkt für Gene identifiziert werden. Die GPRs werden mit Hilfe boolescher Logik modelliert. Interaktionen zwischen dem metabolischen und regulatorischem Netzwerk werden abgebildet, indem die Expression der metabolischen Gene in Abhängigkeit des Zustands des Transkriptionsfaktors als AN bzw. AUS gefolgert werden. Wie für OptReg werden verschiedene Mengen binärer Variablen definiert, um sowohl Knockouts als auch Überexpressionen identifizieren zu können. Die so identifizierten Strategien maximieren die Produktausbeute bei wachstumsoptimalen Bedingungen unter regulatorischen (transkriptionalen) Nebenbedingungen.

FVSEOF

Bei dieser Methode (Flux Variability Scanning based on Enforced Objective Flux) werden Knockout- und Überexpressionskandidaten mit Hilfe einer manipulierten biologischen Zielfunktion bestimmt (Choi et al., 2010). Dabei wird der biologischen Zielfunktion ein konstanter Term für die Ausscheidung des gewünschten Flusses hinzugefügt. Anschließend wird der Anteil dieses Produktflusses schrittweise erhöht und die intrazellulären Reaktionen systematisch nach Reaktionen gescannt, deren Fluss sich im Vergleich zur Ausgangsflusslage obligatorisch erhöhen (für Überexpressionskandidaten) bzw. reduzieren muss (für Knockout-Kandidaten). Eine Erweiterung dieser Methode benutzt das Konzept von Reaktionsgruppierungen (gemeinsam regulierten Reaktionen), um den Lösungsraum einzuschränken (Park et al., 2012). Dabei wird für alle Reaktionen innerhalb einer Gruppe gefordert, dass sie gemeinsam aktiv oder inaktiv sind.

GDLS

Die zuvor beschriebenen Methoden sind grundsätzlich für die Anwendung in genomischen Netzwerken geeignet. Durch den sehr großen Suchraum sind die identifizierten Kandidaten jedoch meist auf drei bis vier Knockouts pro Strategie beschränkt. Diese Limitierung versucht GDLS (Genetic Design through Local Search) zu umgehen, indem ein genetischer Algorithmus mit multiplen lokalen Suchpfaden verwendet wird (Lun et al., 2009). Diese Optimierungsheuristik liefert Lösungen mit deutlich besseren Laufzeiten, kann aber nicht garantieren, dass das globale Optimum identifiziert wird. Eine weitere Beschränkung ist darin zu sehen, dass mit steigendem Umfang der lokalen Suchen die Komplexität der gesamten Optimierungsroutine exponentiell ansteigt.

OptForce

Die Grundidee von OptForce (Ranganathan et al., 2010) ist ähnlich der von FVSEOF. Hier wird sie jedoch als ein zweistufiges Optimierungsproblem formuliert. Es wird zunächst eine Referenzflussverteilung des Wildtyps definiert, der dann eine gewünschte Flussverteilung gegenüber gestellt wird. Zunächst werden alle Reaktionen identifiziert, deren Flüsse sich erhöhen bzw. reduzieren müssen, um die gewünschte Flussverteilung zu erreichen. Anschließend werden alle möglichen Kombinationen minimaler, nicht redundanter Interventionen identifiziert, die obligatorisch zu diesen Änderungen der Flussverteilung führen.

Weitere

Neben den zuvor erläuterten Methoden existieren weitere Ansätze, die im Vergleich zu einigen der diskutierten Methoden entweder Vorteile bei der Berechnung der Interventionen (im Hinblick auf die Laufzeit oder Anzahl identifizierter Strategien, wie zum Beispiel EMILiO (Yang et al., 2011)) oder bei der Qualität der identifizierten Strategien haben (durch die Wahl spezieller Zielfunktionen wie zum Beispiel BiMOMA (Kim et al., 2011)). Deren zu Grunde liegenden Ideen unterscheiden sich jedoch nicht wesentlich von den bereits beschriebenen Methoden, so dass sie hier nicht weiter diskutiert werden. Ausgewählte Methoden werden auch in den Übersichtsartikeln (Zomorodi et al., 2012) und (Tomar und De, 2013) diskutiert.

4.4.2 Ansätze mittels Elementarmodenanalyse

Minimale Metabolische Funktionalität

Verschiedene mathematische Methoden für Metabolic Engineering basieren auf EM. Eine Heuristik zur *in silico* Stammoptimierung, die auf EM zurückgreift, wurde kürzlich von Trinh et al. entwickelt (Trinh et al., 2006, 2008). Ihr Ansatz der minimalen metabolischen Funktionalität (MMF) hat zum Ziel, Organismen zu designen, die außer sehr wenigen verbleibenden Routen zur optimalen Synthese des Produktes oder der Biomasse keine weiteren Pfade etablieren können. Die Vorgehensweise bei diesem Ansatz ist wie folgt:

- Als Ausgangspunkt werden alle EM des Netzwerkes berechnet.
- Anschließend werden für jede Reaktion die Menge der EM, in denen die Reaktion *nicht* vorkommt, bezüglich der maximalen Produktausbeute und der Kardinalität bewertet. Ist in einer solchen EM-Menge noch mindestens ein EM mit maximaler Produktausbeute enthalten, wird die zugehörige Reaktion als Knockout-Kandidat identifiziert. Existieren mehrere Reaktionen, deren Deletion sich nicht negativ auf das Maximum der Produktausbeute auswirkt, so werden die Kandidaten als vorteilhaft betrachtet, bei deren Deletion die meisten EM des Netzwerkes eliminiert werden (bzw. bei denen die Kardinalität der verbleibenden EM-Menge klein ist). Ebenso kann gefordert werden, dass mit den verbleibenden EM weiterhin die Biomassesynthese möglich sein muss.
- Durch mehrmaliges Durchlaufen dieser Vorgehensweise werden iterativ Reaktionen entfernt, bis
 - i) entweder die maximale Ausbeute nicht mehr erhalten werden kann oder
 - ii) die Anzahl an maximal durchführbaren Knockouts erreicht ist.

Mit diesem Ansatz wurden bereits Knockout-Strategien zur Produktion von Ethanol (Trinh et al., 2008) und Carotinoiden (Unrean et al., 2010) mit *E. coli* identifiziert und erfolgreich experimentell validiert.

FluxDesign

Ein alternativer Ansatz, FluxDesign (Melzer et al., 2009), verwendet EM kombiniert mit einer Korrelationsanalyse. Dazu wird zunächst jeder EM bzgl. der Substrataufnahme normiert. Dann wird für jeden EM e_j jede Reaktion i mit dem Kohlenstofffluss der Zielfunktion f (z.B. der Produktausscheidung) verglichen. Ist der Korrelationskoeffizient ρ aller Paare $(e_j, f(e_j))$ deutlich negativ ($\rho \leq -0.7$), so wird Reaktion i als Knockout-Kandidat betrachtet. Ergibt sich ein stark positiver Korrelationskoeffizient ($\rho \geq 0.7$), so wird Reaktion i entsprechend als Überexpressionskandidat bewertet.

4.4.3 Ansatz mittels Minimal Cut Sets

Minimal Cut Sets können natürlicherweise für die Identifikation von Knockouts für Anwendungen des Metabolic Engineering genutzt werden. Mittels der Spezifikation von Target-EM können relevante Designziele wie z.B. die Entfernung aller EM mit einer Produktausbeute von weniger als 90% formuliert werden. Wie in (Klamt, 2006) gezeigt, können zudem komplexe Designziele durch die simultane Berücksichtigung verschiedener Deletionsziele berücksichtigt werden. Die Menge der Zielmoden ergibt sich dann als die Vereinigung der Zielmoden aller zugehörigen Einzelziele. Die Anwendung der MCS für die Identifizierung von Interventionsstrategien ist bisher dadurch limitiert, dass oftmals neben den Target-EM auch viele EM mit gewünschten Funktionalitäten eliminiert werden. Zum Beispiel unterdrücken MCS4 und MCS5 in Abb. 4.3 die suboptimale Synthese von P1 über EM1 und EM3. Allerdings eliminieren diese beiden MCS zusätzlich auch EM1, der zur optimalen Produktion von P1 essentiell ist (Abb. 4.2).

Diese Einschränkung führt zur Grundidee der constrained Minimal Cut Sets, bei denen gewünschte Netzwerkeigenschaften bei der Bestimmung der MCS berücksichtigt werden können. Sie werden in Kapitel 5.2 vorgestellt.

4.4.4 Ansätze mit kinetischer Modellierung

Mit Hilfe dynamischer Netzwerkmodelle können komplexe zelluläre Phänotypen unter verschiedenen Umweltbedingungen oder nach genetischen Eingriffen beschrieben werden. Ein allgemeines Vorgehen zur Entwicklung dynamischer genomweiter metabolischer Netzwerke wurde in (Jamshidi und Palsson, 2008) vorgeschlagen. Derzeit wird dessen Anwendung durch die Nichtverfügbarkeit kinetischer Parameter meist auf kleinformatige metabolische Netzwerke (< 50 Reaktionen) oder Modelle mit vereinfachten Kinetiken begrenzt. Die Aussagekraft solch detaillierter Modelle wird zudem äußerst kritisch diskutiert (von Stockar et al., 2006; Maskow und von Stockar, 2005; Vojinović und von Stockar, 2009; Machado et al., 2012a; Maertens und Vanrolleghem, 2010).

Wegen des Mangels an kinetischen Parametern wird bisher mehrheitlich die strukturelle metabolische Netzwerkmodellierung für stationäre Zustände verwendet. Die Klasse der strukturellen Modellierungsansätze ist bezüglich der verwendeten Informationen und Annahmen deutlich belastbarer. So ist durch die vollständige Sequenzierung des Genoms und der anschließenden Annotierung der Gene das Spektrum der verfügbaren Reaktionen und damit die Netzwerkstruktur eindeutig bekannt. Unsicherheiten ergeben sich lediglich aus noch unbekannten regulatorischen Einschränkungen von bestimmten Reaktionen unter bestimmten Umweltbedingungen.

Dennoch wurden verschiedene Methoden wie z.B. kybernetische Modellierungsansätze und Ensemble Modeling entwickelt, um mit Hilfe geeigneter Annahmen Analysen kinetischer Modelle zu ermöglichen (Song und Ramkrishna, 2012; Kim et al., 2012b; Tran et al., 2008; Young et al., 2008; Wang et al., 2004). Das Prinzip des Ensemble Modeling soll im Folgenden kurz skizziert werden, da es bereits erfolgreich für das Metabolic Engineering verwendet wurde (Rizk und Liao, 2009; Contador et al., 2009).

Ensemble Modeling

Wie oben beschrieben ist die umfassende mechanistische Modellierung des Stoffwechsels (inklusive Regulation und Signaltransduktion) auf Grund unzureichender Informationen der kinetischen Parameter meist nicht zu realisieren. Ein Ansatz, der dennoch eine Annäherung an dieses Ziel versucht, ist durch das Ensemble Modeling gegeben (Tan und Liao, 2012; Tran et al., 2008).

Dabei wird eine Menge dynamischer Modelle generiert, die alle denselben (gemessenen) steady-state erreichen. Die Modelle basieren zudem auf einer identischen Netzwerkstruktur. Sie unterscheiden sich jedoch durch ihre Parametrisierungen wie z.B. durch die Annahme verschiedener Enzym-Expressionslevel. Durch die Verwendung von experimentellen Daten aus Perturbationsexperimenten kann die Menge zulässiger Modelle reduziert werden. Nur diejenigen Teilmengen der konstruierten Modelle, die verschiedene Perturbationen zufriedenstellend wiedergeben können, werden als prädiktiv eingestuft und für die Prädiktion von zukünftigen Perturbationseffekten herangezogen. In einer weiteren Arbeit (Zomorodi et al., 2013) wurde ein Algorithmus vorgeschlagen, um systematisch zu untersuchen, welche Perturbationsexperimente am Besten dazu geeignet sind, die Menge der prädiktiven Modelle zu reduzieren.

Mit Hilfe des Ensemble Modeling-Ansatzes konnten Interventionsziele für die Produktion von aromatischen Kohlenstoffverbindungen (Rizk und Liao, 2009) und Lysin (Contador et al., 2009) in *E. coli* identifiziert werden. Das verwendete Modell umfasste 37 Reaktionen und 34 Metabolite. Dennoch beinhaltet diese Methodik einige Limitierungen. Zum Einen muss der Effekt von Perturbationen *a priori* definiert werden (z.B. eine Überexpression als n -fache Erhöhung der zugehörigen Enzymkonzentration). Des Weiteren muss ein (nur selten verfügbarer) Referenzzustand mit allen Konzentrationen und Flüßsen definiert werden, bzgl. dessen die Auswirkungen analysiert werden.

Tabelle 4.1 fasst alle diskutierten mathematischen Methoden dieses Kapitels zusammen.

Tabelle 4.1: Algorithmen zum Redesign metabolischer Netzwerke. Mögliche Interventionen: Δ Knockout, $\oplus$ Addition heterologer Reaktionen, $\uparrow$ Überexpression, $\downarrow$ Inhibierung

Name	Mathematischer Ansatz	Identifizierbare Interventionen	Referenz
OptKnock	Bilevel, MILP	Δ	(Burgard et al., 2003)
RobustKnock	Multilevel, MILP	Δ	(Tepper und Shlomi, 2010)
OptGene	Genetische Optimierung	Δ	(Patil et al., 2005)
OptStrain	Bilevel, MILP	$\Delta, \oplus$	(Pharkya et al., 2004)
SimOptStrain	Bilevel, MILP	$\Delta, \oplus$	(Kim et al., 2011)
BiMOMA	Bilevel, MINLP	Δ	(Kim et al., 2011)
OptReg	Bilevel, MILP	$\Delta, \uparrow, \downarrow$	(Pharkya und Maranas 2006)
OptOrf	Bilevel, MILP	$\Delta, \uparrow$	(Kim und Reed, 2010)
OptForce	Bilevel, MILP	$\Delta, \uparrow, \downarrow$	(Ranganathan et al., 2010)
EmLiO	Bilevel, MILP	$\Delta, \uparrow, \downarrow$	(Yang et al., 2011)
FSEOV	LP	$\uparrow, \downarrow$	(Choi et al., 2010) (Park et al., 2012)
GDLS	Genetische Optimierung	$\Delta, \uparrow, \downarrow$	(Lun et al., 2009)
MMF	EM	Δ	(Trinh et al., 2006)
FluxDesign	EM und Korrelation	$\Delta, \uparrow, \downarrow$	(Melzer et al., 2009)
MCS	EM und Hitting-Sets	Δ	(Klamt und Gilles, 2004) (Klamt et al., 2006)
Ensemble Modeling	Dynamische Modelle Heuristik	$\Delta, \uparrow, \downarrow$	(Tan und Liao, 2012) (Tran et al., 2008)

5 Neue Methoden zur modellbasierten Identifizierung von Interventionsstrategien

In diesem Kapitel werden neue Methoden und Algorithmen für die modellbasierte Identifizierung von Interventionsstrategien vorgestellt. Die in Kapitel 4.4 vorgestellten existierenden Ansätze optimieren die Flussverteilungen hauptsächlich hinsichtlich der Produktausbeute. In Abschnitt 5.1 wird zunächst CASOP (Computational Approach for Strain Optimization aiming at high Productivity) vorgestellt, ein Ansatz, der explizit sowohl Ausbeute als auch die Produktivität (bzw. als Approximation die Netzwerkkapazität) bei der Identifizierung von Interventionsstrategien berücksichtigt. In Kapitel 5.2 werden die constrained Minimal Cut Sets eingeführt, die eine Erweiterung der ursprünglichen Definition der MCS darstellen. Sie erlauben die Formulierung komplexer Interventionsziele und liefern an Stelle einer speziellen Lösung die Enumeration aller äquivalenten Lösungen des Problems. Zudem ermöglichen sie, mit geeigneten Approximationen, Interventionsstrategien zu identifizieren, deren Ziel entweder auf einer verbesserten Ausbeute oder auf einer erhöhten Produktivität liegt. Sowohl CASOP als auch die constrained MCS sind Ansätze, die auf den EM beruhen. Da deren Berechnung in genomskaligen Netzwerken schwierig oder unmöglich ist, wird in Abschnitt 5.3 der EFMSampler vorgestellt, ein neuer Algorithmus zur Berechnung von EM-Teilmengen in großskaligen Netzwerken. Mit diesem können EM-basierte Methoden für die Anwendung in genomskaligen Netzwerken verfügbar gemacht werden.

5.1 CASOP

5.1.1 Formale Definition

Die meisten constraint-basierten Methoden für das Stammdesign verfolgen als Ziel die Optimierung der Produktausbeute. Jedoch stellt die Maximierung der Biomasse- oder Produktausbeute als Zielfunktional kein allgemein gültiges natürliches Prinzip dar (Schuster et al., 2008). Daher ist es sinnvoll, alternative Designkriterien wie z.B. die spezifische Produktivität (spezifische Produktionsrate des gewünschten Metaboliten) zu betrachten. Diese ist definiert durch:

$$r_P = Y^{P/S} \cdot r_S. \quad (5.1)$$

r_P ist hier die spezifische Produktionsrate [mmol / (gDW · h)], $Y^{P/S}$ die Ausbeute [mol Produkt pro mol Substrat] und r_S die spezifische Substrataufnahmerate [mmol / (gDW·h)].

Da hier ein stöchiometrischer Modellierungsansatz (unabhängig von kinetischen Informationen) verwendet wird, kann die Produktionsrate r_p nicht direkt in einer Zielfunktion für die Optimierung verwendet werden. Aus Gleichung 5.1 wird ersichtlich, dass zum Einen die Ausbeute wichtig für die Erzielung einer hohen Produktionsrate ist. Der zweite Faktor, der die Produktivität bestimmt, ist die spezifische Substrataufnahmerate r_s . Die maximale Substrataufnahmerate $r_{s,max}$ ist die Menge an Substrat, die in einer bestimmten Zeitspanne verstoffwechselt werden kann. Damit kann diese Rate auch als Maß für die Netzwerkkapazität (Durchsatz) interpretiert werden.

Ohne kinetische Daten kann die Kapazität eines metabolischen Netzwerks unter gegebenen Umweltbedingungen nicht mit verlässlicher Sicherheit quantifiziert werden. Dennoch ist es mit stöchiometrischen Ansätzen möglich, mit Hilfe vereinfachender Annahmen Aussagen bezüglich des Beitrags einzelner Reaktionen oder Pfade zur Netzwerkkapazität vorzunehmen. So beschränkt zum Beispiel die langsamste essentielle Reaktion für eine gewünschte Funktionalität dessen Gesamtumsatz. Eine nützliche Approximation der Flexibilität ist z.B. die Häufigkeit des Auftretens einer Reaktion in den zur Verfügung stehenden oder gewünschten Routen (EM). Gibt es z.B. mehrere parallele Pfade, die die Umsetzung eines Substrats in ein Produkt ermöglichen, so kann die Gesamtrate sich potentiell aus der Summe der Geschwindigkeiten der parallelen Pfade zusammensetzen. Wenn eine Reaktion in vielen dieser Pfade vorkommt, sollte unter Umständen vermieden werden, diese zu eliminieren, auch wenn einige dieser Pfade nur eine geringe Ausbeute ermöglichen. Auf dem Ansatz der Approximation der potentiellen Netzwerkkapazität durch die Netzwerkflexibilität basiert auch die CASOP-Methodik.

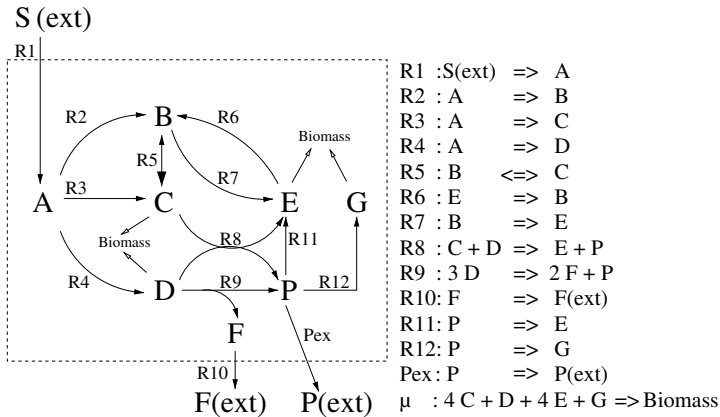


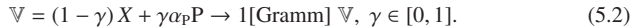
Abbildung 5.1: Beispielnetzwerk N_2

Anhand des Beispielnetzwerkes N_2 (Abbildung 5.1) wird die Methode erläutert. Das Netzwerk besteht aus 14 Reaktionen und acht internen Metaboliten, deren molare Masse einheitlich als 0.1 g / mol definiert ist. Das Substrat S kann entweder zu den Produkten F und P oder zur Biomasse X umgesetzt werden. Die Stöchiometrie der Biomasse-Synthese Reaktion μ ist so gewählt, dass 1 mol X einem Gramm Biotrockenmasse entspricht. Für

dieses Netzwerk sollen im Folgenden Interventionsstrategien identifiziert werden, die zu einer gesteigerten Ausbeute und gesteigerten Ausscheiderate des Metaboliten P führen.

Der Metabolit P kann sowohl durch Reaktion R8 als auch mit Reaktion R9 gebildet werden. Nur Routen, die Reaktion R8 verwenden, ermöglichen die optimale Ausbeute von 1 mol P pro mol Substrat S (wenn zudem keine Biomasse gebildet wird). Die Pfade, die Reaktion R9 verwenden, ermöglichen hingegen eine deutlich geringere Ausbeute von 1/3, da R9 drei D verbraucht. Eine Verbesserung der Ausbeute $Y^{P/S}$ kann demnach leicht erreicht werden, indem Reaktion R9 deletiert wird. Da Ausbeute $Y^{P/S}$ und Kapazität r_S nicht unabhängig voneinander optimiert werden können, kann dennoch nicht gefolgert werden, dass der Knockout von R9 auch zwingend zu einer höheren Produktivität bezüglich P führt. Mit Hilfe üblicher Reaktionskinetiken (wie z.B. Michaelis-Menten) könnte man leicht ein Modell definieren, für das der Knockout von R9 sogar zu einer verringerten Ausscheiderate von P führt. Für die Optimierung der Produktivität muss stets zwischen Ausbeute auf der Einen und Kapazität bzw. Flexibilität auf der anderen Seite abgewogen werden.

In praxisrelevanten Anwendungen wird als zusätzliches Designkriterium gefordert, dass der Organismus weiterhin wachsen kann. Für das Beispielnetzwerk bedeutet dies, dass es weiterhin möglich sein muss, die Präkursoren der Biomasse C, D, E und G bilden zu können. Besonders interessant sind mithin solche Flusslagen, bei denen in einem bestimmten Verhältnis gleichzeitig Biomasse und Produkt gebildet werden. Um alle Möglichkeiten der simultanen Produktion von X und P abzubilden, wird das Netzwerk zunächst leicht modifiziert. Dafür wird die Biomasse vorübergehend als interner Metabolit betrachtet. Zudem wird eine künstliche Reaktion R_V definiert, bei der aus den Edukten P und X ein Hilfsmetabolit V gebildet wird. Dieser Hilfsmetabolit V dient als externer „Container“, dessen anteilige Zusammensetzung von P und X durch die Stöchiometrie von R_V definiert wird:



Der Parameter γ weist den Anteil der Masse von P an V zu und bestimmt das Verhältnis der Synthese von P und Biomasse. Der Fall $\gamma = 0$ entspricht der exklusiven Biomassesynthese ohne Produktbildung, während für $\gamma = 1$ die exklusive Produktion des gewünschten Metaboliten P erfolgt. Um Flussverteilungen für verschiedene Verhältnisse von Wachstum und Produktsynthese zu vergleichen, kann γ zwischen 0 und 1 variiert werden. Jedem konkreten γ entspricht ein Anteilsszenario S_γ , für das alle EM berechnet werden. Um die stöchiometrische Konsistenz in Gleichung 5.2 zu erhalten, wird P mit dem Kehrwert seiner Molmasse, $\alpha_P[\text{mol/g}]$, gewichtet. Für das Beispielnetzwerk folgt, dass $\alpha_P = 10\text{mol/g}$.

Zur Veranschaulichung der Methodik wird γ nun in diskreten Schritten von 0 bis 0.9 mit einer Schrittweite von 0.1 erhöht. Der größte Wert für γ wird hier bewusst kleiner als 1 gewählt, da die gekoppelte Produkt- und Biomassesynthese betrachtet werden soll und nicht die exklusive Produktion von P (wie im Fall $\gamma = 1$). Für jedes der zehn resultierenden Szenarien S_γ werden die EM berechnet und alle EM mit strikt positiver Ausbeute $Y^{V/S}$ werden für weitere Betrachtungen selektiert. Durch die implizite Definition der Produktsynthese in V kann die molare Ausbeute von gebildetem P aus der Ausbeute von V berechnet werden: $Y^{P/S} = \gamma\alpha_P Y^{V/S}$. Durch die Verschiebung der Stöchiometrie des künstlichen Hilfsmetaboliten V von der reinen Biomassesynthese hin zu gekoppelter Produkt- und Biomassesynthese (mit steigendem Anteil von P an V) verändern sich auch die berechneten EM. Bei vielen wird der Support unverändert bleiben und sich nur die konkreten

Werte der jeweiligen Flussverteilung verändern. Es können bestimmte Supportmuster einiger EM auch komplett verschwinden oder andere neu erscheinen (Abbildung 5.2). Es sind diese Veränderungen im Spektrum aller EM, die im Folgenden analysiert werden, um geeignete Eingriffsmaßnahmen zu identifizieren.

Für ein gegebenes Anteilsszenario $\mathbb{S}_\gamma$ sind die EM mit strikt positiver Ausbeute $Y^{V/S}$ durch $n(\gamma)$ bestimmt. Diese EM sollen nun bezüglich verschiedener Kriterien gewichtet werden. Wenn keine weiteren Annahmen oder Wertungen vorgenommen werden, können z.B. alle EM zunächst als gleich wahrscheinlich bzw. gleich relevant angenommen werden. Falls man jedoch das biologische Ziel der Zelle bzw. das ökonomische Ziel des Modellierers berücksichtigen möchte, so sollten EM mit einer hohen Ausbeute $Y^{V/S}$ ein größeres Gewicht bekommen als EM mit einer kleineren Ausbeute. Eine Definition eines Gewichts $v_{i,\gamma}$ für die EM $_i^r$ ($i = 1, \dots, n(\gamma)$) eines Szenarios $\mathbb{S}_\gamma$, das die erwähnten Anforderungen erfüllt, ist gegeben durch:

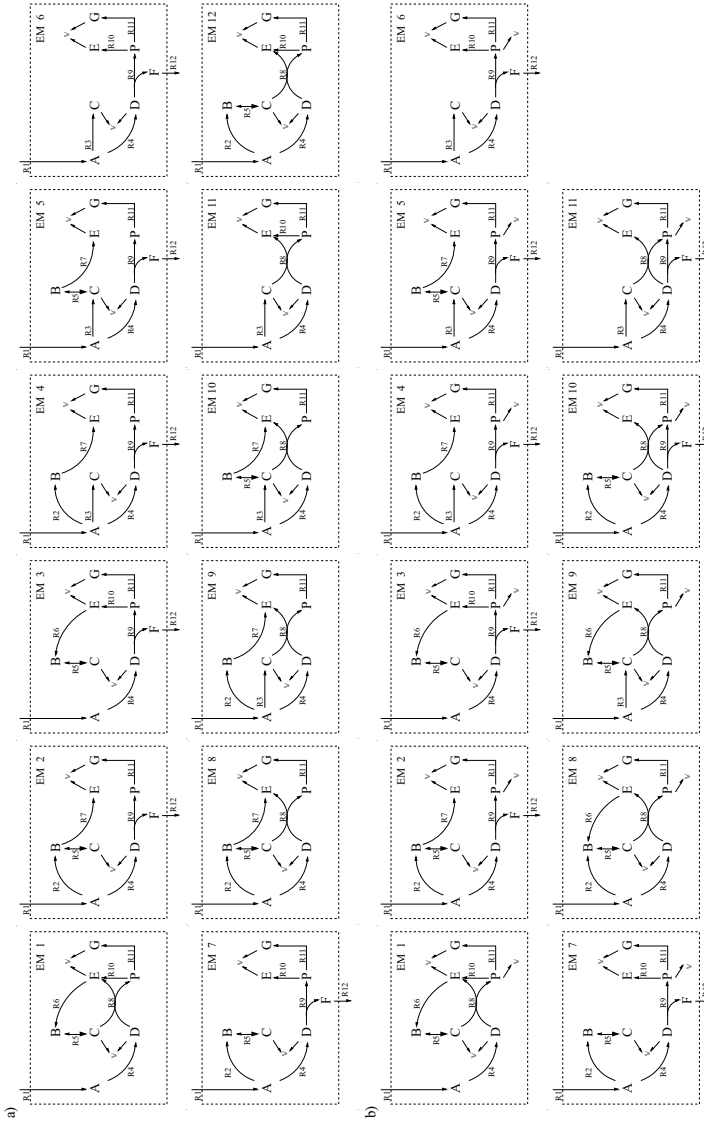
$$v_{i,\gamma} = \frac{\left(Y_i^{V/S}\right)^k}{\sum_{i=1}^{n(\gamma)} \left(Y_i^{V/S}\right)^k}. \quad (5.3)$$

Mit Hilfe des Parameters $k \geq 0$ kann die Bedeutung der Ausbeute für das Gewicht beeinflusst werden. Wird $k = 0$ gewählt, so bekommt jeder EM das gleiche Gewicht, unabhängig von der spezifischen Ausbeute der einzelnen EM. Kleine Werte für k führen nur zu geringen Unterschieden in der Gewichtung der EM, so dass suboptimale EM (in Bezug auf Ausbeute) noch immer entsprechend berücksichtigt werden. Mit steigendem k wird die Ausbeute der EM das Gewicht immer stärker beeinflussen, wobei eine hohe Ausbeute sich auch in einem hohen Gewicht für den entsprechenden EM niederschlägt.

Die Gewichte der EM werden nun verwendet, um ein Maß für die Relevanz $\omega_\gamma(r_j)$ für jede Reaktion r_j zu bestimmen (pro betrachtetes Szenario $\mathbb{S}_\gamma$). Es wird definiert als die Summe aller EM-Gewichte, in denen die Reaktion r_j enthalten ist (Index h läuft über alle EM, in denen die Reaktionsrate $r_j \neq 0$, ist):

$$\omega_\gamma(r_j) = \sum_{h: EM_h^r(r_j) \neq 0} v_{h,\gamma}. \quad (5.4)$$

Gleichung 5.4 zeigt zusammen mit Gleichung 5.3 eine weitere Bedeutung des Parameters k . In Abhängigkeit des gewählten k kann die Bedeutung dieses Maßes zwischen zwei Ausprägungen justiert werden. Für $k = 0$ ist die Relevanz $\omega(r_j)$ identisch zur relativen Auftrittshäufigkeit der Reaktionen. Somit schätzt ω in diesem Fall den Beitrag einer Reaktion für die Flexibilität bzw. Produktivität. Für große k (≥ 2) wird mit ω hingegen die Relevanz einer Reaktion für die Erzielung optimaler Produktausbeuten bestimmt. Grundsätzlich kann $k \geq 0$ gewählt werden. Es hat sich jedoch gezeigt, dass ganzzahlige $0 \leq k \leq 10$ genügen, um ω sowohl als Maß für die Relevanz einer Reaktion für die Produktausbeute als auch als Maß für die Flexibilität bzw. Produktivität verwenden zu können.

Abbildung 5.2: Elementarmoden des Netzwerkes N_2 für a) $\gamma = 0$, b) $\gamma = 0.9$.

Eine einfache Umformung von Gleichung 5.4 führt zu einer äquivalenten Darstellung von $\omega_\gamma(r_j)$:

$$\omega_\gamma(r_j) = \sum_{h:EM_h^r(r_j) \neq 0} v_{h,\gamma} = \frac{\sum_{h:EM_h^r(r_j) \neq 0} (Y_h^{\nabla/S})^k}{\sum_{h=1}^{n(\gamma)} (Y_h^{\nabla/S})^k} = \frac{m_j(\gamma)}{n(\gamma)} \cdot \frac{(\overline{Y_j^{\nabla/S}})^k}{(\overline{Y^{\nabla/S}})^k} \quad (5.5)$$

wobei $m_j(\gamma)$ die Anzahl an EM bestimmt, an denen Reaktion r_j in $\mathbb{S}_\gamma$ teilnimmt. Gleichung 5.5 zeigt, wie das Relevanzmaß die Beiträge einer Reaktion zur Flexibilität und zur Ausbeute erfasst. Die relative Auftrittshäufigkeit $\frac{m_j(\gamma)}{n(\gamma)}$ ist unabhängig vom Parameter k und spiegelt den Beitrag der Reaktion r_j an der Flexibilität des Netzes wider. Der zweite Faktor berücksichtigt die gemittelte (und evtl. gewichtete) Ausbeute aller EM, in denen Reaktion r_j auftritt $\left(\overline{Y_j^{\nabla/S}}\right)^k$, normalisiert zur durchschnittlichen (evtl. gewichteten) Ausbeute aller EM $\left(\overline{Y^{\nabla/S}}\right)^k$ in Szenario $\mathbb{S}_\gamma$. Dieser Term reflektiert das Ausbeutekriterium und sein Beitrag am Gewicht kann durch den Parameter k beeinflusst werden.

Die Werte des Relevanzmaßes $\omega(r_j)$ sind so definiert, dass sie im Intervall $[0, 1]$ liegen. Essentielle Reaktionen erhalten eine Relevanz von 1 (unabhängig von k), während Reaktionen, die in keinem ∇ -produzierenden EM enthalten sind, eine Relevanz von 0 zugewiesen wird. Der Ansatz der Gewichtung von EM und anschließender Bewertung der Relevanz von Reaktionen wurde bereits für andere Anwendungen verwendet, wie z.B. den „control-effective fluxes“ (Stelling et al., 2002). Für die hier vorgestellte Anwendung zur Stammoptimierung wurden jedoch neue Gewichte, Bewertungsmaße und Bezeichnungen definiert und deren Interpretation entsprechend angepasst.

Im Beispielnetzwerk N_2 wurden die Relevanzen für $k = 0, 2$ und 10 berechnet (Abbildung 5.3). Wie beschrieben ist die Synthese des Metaboliten P implizit in der Synthese von ∇ für $\gamma > 0$ enthalten. Die maximal erreichbare Ausbeute von P auf dem Substrat S ist $Y^{P/S} = 1$. Die Verwendung der Reaktionen R1, R4 und R9 hat eine geringe Ausbeute von $Y^{P/S} = 1/3$ zur Folge. Es gibt vier essentielle Reaktionen für Biomasse- und somit für ∇ -Synthese (R1, R4, R12 und trivialerweise μ), denen jeweils eine Relevanz von 1 zugewiesen wird (unabhängig vom gewählten Anteilsszenario S_γ und Parameter k (siehe Abbildung 5.3)).

Die Verläufe der Relevanzen der übrigen Reaktionen sind differenzierter. R2, R3 und R11 beginnen bei einer mittleren Relevanz und zeigen eine leicht abnehmende Tendenz für größere γ . Die Relevanz dieser drei Reaktionen ist nahezu konstant für verschiedene Werte von k . Die Relevanz der Reaktion R8 bleibt annähernd konstant mit zunehmendem γ . Jedoch ist für größere k (Schwerpunkt auf der Ausbeuteoptimalität der Flussverteilungen) eine deutlich höhere Relevanz zu beobachten. Dies kann dadurch erklärt werden, dass alle optimalen Routen für P-Synthese sowie für die Synthese der Biomassekomponenten E und G eine Beteiligung von R8 erfordern. Ein anderes Verhalten kann für R5, R6 und R7 beobachtet werden: Mit steigendem γ wird den Reaktionen R5 und R6 eine höhere und R7 eine geringere Relevanz zugeordnet. Da das Produktionsverhältnis von Metabolit E und Produkt P in Richtung P-Synthese verschoben wird, muss eine größere Menge von E über R5 und R6 recycelt werden, um Reaktion R8 betreiben zu können. Reaktion R7 wird aus dem gleichen Grund mit wachsendem γ entbehrlicher. Dieser Effekt

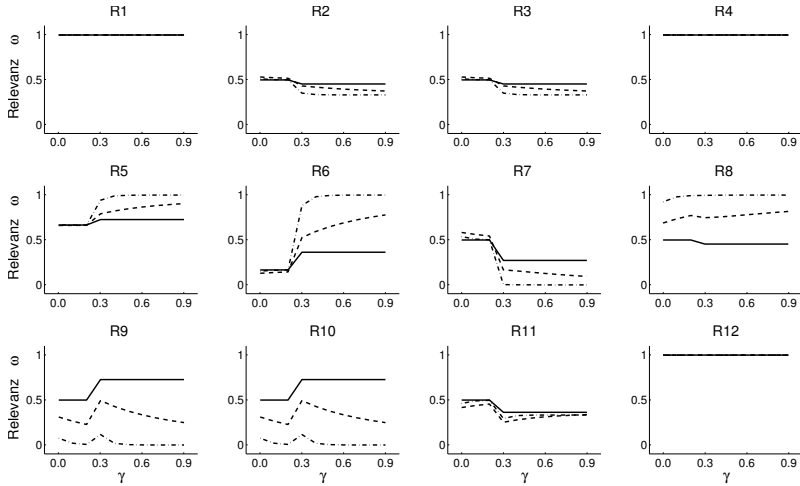


Abbildung 5.3: Relevanz der Reaktionen des Netzwerkes N_2 in Abhängigkeit von γ für $k = 0$ (Linie), 2 (Strichlinie), 10 (Strich-Punkt-Linie).

wird bei größeren k noch erhöht, da das Recycling von E und der anschließende Umsatz über R8 eine optimale Ausbeute gewährleistet.

Die Relevanzen der Reaktionen R9 und R10 zeigen ein zueinander identisches Verhalten, da beide Reaktionen gekoppelt sind und daher ein Enzym- oder Reaktions-Subset bilden (Pfeiffer et al., 1999). Für $k = 0$ werden die Relevanzen größer, weil über R9 und R10 P gebildet werden kann, welches in zunehmenden Mengen mit steigendem γ erforderlich ist. Da jedoch bei der Verwendung von R9 und R10 nur eine geringe Ausbeute erzielt werden kann, wird diesen Reaktionen eine geringere Relevanz beigemessen, wenn für k größere Werte verwendet werden. Zudem zeigt der Verlauf der Relevanzen von R9 und R10 auch, dass es nicht-monotone Abhängigkeiten von γ geben kann, insbesondere wenn bestimmte Supportmuster von EM verschwinden oder neu erscheinen (z.B beim Schritt von $\gamma = 0.2$ auf 0.3 , Abbildung 5.3).

Die Relevanzen werden nun genutzt, um eine Bewertung des Interventionspotentials vorzunehmen und eine Rangliste geeigneter Knockout- und Überexpressionskandidaten aufzustellen. Dafür sind verschiedene Kriterien denkbar, von denen im Folgenden zwei detailliert diskutiert werden. Eine einfache Möglichkeit, das Interventionspotential zu bewerten, ist gegeben durch die Relevanz $\omega(r_j)$ einer Reaktion r_j für ein Szenario $\mathbb{S}_\gamma$ mit hohem Anteil des Produktes am externen Container $\mathbb{V}$; also z.B. für $\gamma = 0.9$. Das Interventionspotential Z_1 ist dann definiert als:

$$Z_1(r_j) = \omega_{0.9}(r_j). \quad (5.6)$$

Z_1 kann Werte zwischen 0 und 1 annehmen, wobei kleine Werte auf Knockout-Kandidaten und große Werte (≈ 1) auf Überexpressionskandidaten hinweisen. Im Beispielnetzwerk N_2 führt die Verwendung von Z_1 zu folgenden Interventionskandidaten: Mit einer ausbeu-

getriebenen Strategie ($k = 10$) werden den Reaktionen R7, R9 und R10 die kleinsten Bewertungen zugeordnet (≈ 0) woraufhin sie als Knockout-Kandidaten identifiziert werden (Tabelle 5.1). Den essentiellen Reaktionen R1, R4 und R12 sowie den Reaktionen R5, R6 und R8 werden hohe Z_1 -Werte zugewiesen (≈ 1) und sie damit als Überexpressionskandidaten gewertet. Wenn hingegen eine Strategie mit Fokus auf Netzwerkflexibilität gewünscht ist, so gibt es zum Einen weniger Knockout-Kandidaten und zum Anderen werden die Reaktionen R9 und R10 nun sogar als Überexpressionskandidaten identifiziert (Tabelle 5.1).

Tabelle 5.1: Interventionspotential Z_1 und Z_2 der Reaktionen des Beispielnetzwerkes für verschiedene k . Z_1 entspricht den Relevanzen der Reaktionen für $\gamma = 0.9$.

Reaktion	Z_1			Z_2		
	$k = 0$	$k = 2$	$k = 10$	$k = 0$	$k = 2$	$k = 10$
R1	1.00	1.00	1.00	0.00	0.00	0.00
R2	0.45	0.38	0.33	-0.05	-0.16	-0.18
R3	0.45	0.38	0.33	-0.05	-0.16	-0.18
R4	1.00	1.00	1.00	0.00	0.00	0.00
R5	0.73	0.91	≈ 1.00	0.06	0.24	0.33
R6	0.36	0.78	≈ 1.00	0.20	0.65	0.85
R7	0.27	0.09	≈ 0.00	-0.23	-0.49	-0.54
R8	0.45	0.82	≈ 1.00	-0.05	0.13	0.07
R9	0.73	0.25	≈ 0.00	0.23	-0.06	-0.07
R10	0.73	0.25	≈ 0.00	0.23	-0.06	-0.07
R11	0.36	0.34	0.33	-0.14	-0.08	-0.13
R12	1.00	1.00	1.00	0.00	0.00	0.00

Ein alternatives Maß für das Interventionspotential berücksichtigt den Verlauf der Relevanzen in Abhängigkeit von γ . Da Mikroorganismen oftmals optimal in Bezug auf die Ausbeute ihrer Biomasse wachsen (Ibarra et al., 2002) kann das Wissen über das Wildtypverhalten als Referenzverteilung für die Relevanzen der Reaktionen berücksichtigt werden. Für wachstumsoptimierte Wildtypstämme kann angenommen werden, dass die Flüsse (zumindest qualitativ) entsprechend den Relevanzen für $\gamma = 0$ verteilt sind. Für produktionsoptimierte Stämme sollte dann analog gelten, dass die Flüsse entsprechend den Relevanzen bei hohem γ (≥ 0.9) verteilt sind. Die Differenz der Relevanzen der Reaktionen zwischen diesen verschiedenen Umweltbedingungen führt zur Definition eines alternativen Maßes des Interventionspotentials:

$$Z_2(r_j) = \omega_{0.9}(r_j) - \omega_0(r_j). \quad (5.7)$$

Die Werte für Z_2 liegen im Intervall $[-1, 1]$, wobei negative Werte auf Knockout- und positive Werte auf Überexpressionskandidaten verweisen. Wird Z_2 als Maß des Interventionspotentials für das Beispielnetzwerk verwendet, so ergibt sich teilweise eine andere Rangfolge (im Vergleich zu Z_1) der Knockout- und Überexpressionskandidaten (Tabelle 5.1). Da R1, R4 und R12 im Wildtyp essentiell sind, werden diese Reaktionen bereits

aktiv sein (mit hoher Genexpression für die entsprechenden Enzyme). Es gibt somit keine Notwendigkeit, an diesen Punkten einzugreifen. Das Interventionspotential Z_2 weist diese Reaktionen demnach nicht als Überexpressionskandidaten aus. Ähnliche Argumente gelten für R8. In Übereinstimmung mit Z_1 werden für eine ausbeuteoptimale Strategie R5 und R6 als Überexpressionkandidaten und R7 als Knockout-Kandidat identifiziert. R9 und R10 werden auch durch Z_2 als Überexpressionskandidaten identifiziert, wenn der Fokus auf der Flexibilität bzw. Produktivität liegt. Im Gegensatz zur Bewertung des Interventionspotentials mit Z_1 hat ein Knockout von R9 und R10 nur eine untergeordnete Bedeutung, wenn eine ausbeuteoptimale Strategie (großes k) verfolgt wird. Auch hier ist die Erklärung darin zu sehen, dass mittels der Relevanzen für $\gamma = 0$ abgeschätzt werden kann, dass diese Reaktionen bereits im Wildtyp geringe Aktivitäten haben sollten und daher nicht zwingend ausgeschaltet werden müssen.

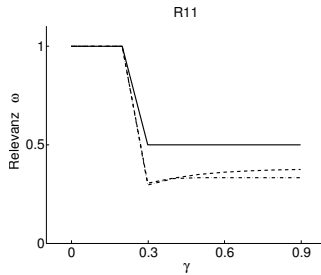


Abbildung 5.4: Relevanz von R11 nach dem Knockout von R7 für $k = 0, 2, 10$ (Linie, Strichlinie, Strich-Punkt-Linie).

Die Interventionspotentiale Z_1 und Z_2 können teilweise zu verschiedenen Ergebnissen führen, wobei die Verwendung von Z_2 in den bisherigen Anwendungen favorisiert wurde. Identische Ziele werden dann identifiziert, wenn $\omega_{0.9}(r_j)$ groß ist und $\omega_0(r_j)$ klein ist und umgekehrt. Wie oben beschrieben, können die Maße Z_1 und Z_2 unmittelbar verwendet werden, um einzelne Knockout- oder Überexpressionskandidaten zu identifizieren. Wenn mehrere Knockouts durchgeführt werden sollen, kann wie folgt vorgegangen werden:

Für das Beispielnetzwerk N_2 wurde R7 als bester Knockout-Kandidat identifiziert (für alle k und für Z_1 und Z_2). Um einen zweiten Knockout-Kandidaten zu finden, werden die EM, die R7 enthalten, ausgeschlossen und die Relevanzen sowie die Interventionspotentiale neu berechnet. Im Beispiel sind die Relevanzen in der zweiten Iteration mit Ausnahme von R11 nahezu unverändert (nicht gezeigt). Nach dem Knockout von R7 wird Reaktion R11 für $\gamma < 0.3$ essentiell, während für $\gamma \geq 0.3$ die Relevanz von R11 auf nicht essentielle Werte abfällt (Abbildung 5.4). Der zusätzliche Knockout von R11 würde somit implizieren, dass keine Netzwerkszenarien mit $\gamma < 0.3$ mehr existieren. Dies ist ein wünschenswertes Ergebnis, da nun Biomasse- und Produktsynthese gekoppelt sind und unter Wachstumsbedingungen mindestens 30% des Substrates für die Herstellung des Metaboliten P umgeleitet werden muss. Das kleinste γ , für das noch EM existieren, gibt somit den Mindestmassenanteil des Produktes an der gekoppelten Ausbeute von Produkt und Biomasse an.

Obwohl nicht explizit im CASOP-Ansatz gefordert, können die identifizierten Ziele

somit zu obligatorischer Kopplung führen. Zugleich zeigt dieses Beispiel den Vorteil der iterativen Berechnung der Interventionskandidaten. In der ersten Iteration war Reaktion R11 nicht automatisch als der zweitbeste Knockout-Kandidat erkennbar. Der zusätzliche Knockout von R2 oder R3 (den besseren Knockout-Kandidaten der ersten Iteration) führt jedoch nicht zu obligatorischer Kopplung.

Die Berücksichtigung von identifizierten Überexpressionskandidaten in nachfolgenden Iterationen ist schwierig, da eine quantitative Vorhersage, wie die spezifischen Durchsätze der EM sich ändern werden, nicht möglich ist. Ein Ansatz wäre, in nachfolgenden Iterationen nur die EM zu betrachten, in denen die entsprechende Reaktionsgeschwindigkeit ungleich Null ist. Wenn Kombinationen von Knockouts und Überexpressionen bestimmt werden sollen, kann man zunächst die gewünschte Anzahl von Knockout-Kandidaten identifizieren (die durch Entfernen der zugehörigen Reaktion einfach berücksichtigt werden können) und anschließend nach möglichen Überexpressionskandidaten suchen. Generell wird die Prozedur gestoppt, wenn die gewünschte Anzahl von Eingriffen erreicht ist oder wenn keine weitere nichtessentielle Reaktion vorhanden ist. Folgender Pseudocode fasst die CASOP-Methode zusammen:

-
1. Definition des artifiziellen Containers $\mathbb{V}$ (zusammengesetzt aus Biomasse X und dem gewünschten Produkt P).
 2. Schrittweise Erhöhung des Anteils (γ) von P an $\mathbb{V}$ und Berechnung der EM für jedes Szenario $\mathbb{S}_\gamma$ (grundsätzlich genügen die EM für ein bzw. zwei $\mathbb{S}_\gamma$, um Z_1 bzw. Z_2 zu berechnen).
 3. Auswahl der gültigen EM (bezogen auf Umweltnebenbedingungen oder bereits berücksichtigte Interventionen).
 4. Wahl eines geeigneten k (Justierung des Verhältnisses zwischen Flexibilität und Ausbeuteoptimalität) sowie Berechnung der EM-Gewichte ($v_{i,\gamma}$) und Relevanzen ($\omega_\gamma(r_j)$) der Reaktionen (für jedes $\mathbb{S}_\gamma$).
 5. Berechnung der Interventionspotentiale Z_1 und bzw. oder Z_2 und Identifizierung der Knockout- und bzw. oder Überexpressionskandidaten.
 6. Wenn die Anzahl gewünschter Interventionen erreicht ist $\rightarrow$ Stop (sonst weiter mit Schritt 3).
-

5.1.2 Erweiterungen

Die CASOP-Methodik kann flexibel angepasst werden, um spezifische Anforderungen zu erfüllen. So sind alternative Maße des Interventionspotentials wie zum Beispiel der durchschnittliche Anstieg bzw. Abstieg der Relevanzen der Reaktionen denkbar. Je nachdem, ob der Verlauf der Relevanzen in Abhängigkeit von γ einen positiven oder negativen Anstieg ausweist, werden die zugehörigen Reaktionen als Überexpressions- oder Knockout-Kandidaten klassifiziert. Mit diesem Maß kann vermieden werden, dass Reaktionen mit nichtmonotonem (und schwierig zu interpretierendem) Verlauf der Relevanzen als Interventionskandidaten vorgeschlagen werden. Im Folgenden werden zwei weitere Anpassungen bzw. Erweiterungen der CASOP-Methodik vorgestellt. Zum Einen wird erläutert, wie mit Hilfe eines alternativen Elementarmodenmaßes thermodynamische Informationen berücksichtigt werden können. Des Weiteren wird skizziert, wie der Einfluss von Co-Faktoren für die Produktsynthese abgeschätzt werden kann.

Berücksichtigung thermodynamischer Informationen

Thermodynamische Informationen können in der CASOP-Methodik berücksichtigt werden, indem alternative Gewichte für die EM definiert werden.

Eine Möglichkeit, die Kapazität bzw. den potentiellen Durchsatz eines EM zu approximieren, ist durch die freie Gibbs'sche Energie der Gesamtreaktion (der Nettostöchiometrie) eines EM gegeben. Im einfachsten Fall wird die Gibbs'sche Energie eines EM unter Standardbedingungen gemäß

$$\Delta_{P_i} G^0 = \sum G_f^0(\text{Produkte}) - \sum G_f^0(\text{Edukte}) \quad (5.8)$$

berechnet, wobei G_f^0 den Standardbildungsenthalpien der Metabolite entsprechen. Die Edukte und Produkte der Gesamtreaktion sind gewöhnlich externe Metabolite. Ist man daran interessiert, die EM ausschließlich im Hinblick auf ihre Kapazität zu gewichten, so kann in Gleichung 5.3 das Gewicht durch

$$v_{i,\gamma} = \frac{(-\Delta_{P_i} G^0)}{\sum_{i=1}^{n(\gamma)} (-\Delta_{P_i} G^0)} \quad (5.9)$$

ersetzt werden. Um die Eigenschaft zu erhalten, dass die $\omega_\gamma(r_j)$, die sich aus diesem EM-Gewicht ergeben, im Intervall $[0, 1]$ liegen, muss gewährleistet sein, dass alle $\Delta_{P_i} G^0$ negativ sind. Deshalb werden alle EM, deren $\Delta_{P_i} G^0$ nicht-negativ ist, vor der Berechnung der EM-Gewichte verworfen. Dadurch wird sogleich garantiert, dass nur thermodynamisch zulässige (unter Standardbedingungen) EM verwendet werden. Die $\Delta_{P_i} G^0$ stellen allerdings nur eine Approximation der, von den Konzentrationen der Metabolite abhängigen, *in vivo* realisierten $\Delta_{P_i} G$ dar. Da in realen Netzwerken jedoch meist nur sehr wenige EM positive $\Delta_{P_i} G^0$ aufweisen, ist der Fehler, der durch evtl. zu Unrecht verworfene EM entsteht, vernachlässigbar. Zudem können, wenn extrazelluläre Konzentrationsdaten eines stationären Zustandes bekannt sind, die zugehörigen $\Delta_{P_i} G$ an Stelle der $\Delta_{P_i} G^0$ verwendet werden.

Für die EM-Gewichte in Gleichung 5.9 und den resultierenden Relevanzen $\omega_y(r_j)$ ist der Parameter k zunächst obsolet, da nicht direkt zwischen Kapazität und Ausbeute abgewogen werden kann. Soll diese Möglichkeit weiterhin erhalten bleiben, so kann ein alternatives Gewicht definiert werden, in dem sowohl die Ausbeute als auch die $\Delta_{P_i} G^0$ explizit berücksichtigt werden:

$$v_{i,y} = \frac{\frac{(Y_i^{y/s})^{k_1}}{\sum_{i=1}^{n(y)} (Y_i^{y/s})^{k_1}} \cdot \frac{(-\Delta_{P_i} G^0)^{k_2}}{\sum_{i=1}^{n(y)} (-\Delta_{P_i} G^0)^{k_2}}}{\sum_{i=1}^{n(y)} \frac{(Y_i^{y/s})^{k_1}}{\sum_{i=1}^{n(y)} (Y_i^{y/s})^{k_1}} \cdot \frac{(-\Delta_{P_i} G^0)^{k_2}}{\sum_{i=1}^{n(y)} (-\Delta_{P_i} G^0)^{k_2}}} \quad (5.10)$$

Für dieses EM-Gewicht kann mit Hilfe der beiden Parameter k_1 und k_2 zwischen der Ausbeute- ($k_1 > k_2$) und Kapazität ($k_2 > k_1$) abgewogen werden. Wie zuvor können die resultierenden Relevanzen der Reaktionen nur Werte zwischen 0 und 1 annehmen (wenn alle $\Delta_{P_i} G^0 \leq 0$).

Analyse von Co-Faktoren und Co-Metaboliten

Eine weitere nützliche Anwendung der CASOP-Heuristik ist die Bewertung des Bedarfs an Co-Faktoren (wie z.B. NAD/NADH oder ADP/ATP) bzw. Co-Metaboliten (z.B. CO_2) für die Produktsynthese. Dafür werden zunächst künstliche „Hilfsreaktionen“ definiert, die als Quelle bzw. Senke für die entsprechenden Co-Faktoren dienen. Für die erweiterten Netzwerke werden anschließend die EM für verschiedene Szenarien $\mathbb{S}_y$ berechnet sowie die benötigten Gewichte und Maße der neu eingeführten Reaktionen bestimmt.

In den erweiterten Netzwerken ändern sich die Mengen der EM und damit auch die CASOP-Maße der anderen Reaktionen. Daher sollten sie nur zur Bewertung der artifiziellen Reaktionen verwendet werden. Wenn inhaltlich nicht anders begründet (wie im Fall von CO_2), sollte die Aufnahme bzw. Ausscheidung der Co-Faktoren separat untersucht werden, da die simultane Verfügbarkeit der entsprechenden Quellen und Senken zu unterschiedlichen Ergebnissen führen kann. Konkrete Beispiele der Co-Faktor-Bewertung werden später in Kapitel 6.2.1 vorgestellt.

Grundsätzlich lässt sich mit diesem Vorgehen auch für jeden anderen Metaboliten die Relevanz einer zusätzlichen Quelle bzw. Senke analysieren. Es muss wie zuvor lediglich das Netzwerk um eine artifizielle Aufnahme- bzw. Ausscheidereaktion für den entsprechenden Metaboliten erweitert werden. Sofern es sich dabei um verstoffwechselbare Metabolite handelt, sollte die Ausbeute des Produktes dann bezüglich der gemeinsamen Aufnahme des Substrates und des betrachteten Metaboliten berechnet werden.

Implementierung

Die CASOP-Funktionalität wurde in das *CellNetAnalyzer*-Paket integriert (Klamt et al., 2007). Diese Matlab-Toolbox kann sowohl kommandozeilengesteuert (mittels einer API Schnittstelle (Klamt und von Kamp, 2011)) als auch über eine graphische Benutzeroberfläche bedient werden. Zur Berechnung der EM sind zudem die Programme Metatool (von Kamp und Schuster, 2006) und EFMtool (Terzer und Stelling, 2008) integriert. Neben einer umfangreichen Funktionalität zur Elementarmodenanalyse steht das gesamte Spektrum an Methoden zur stationären Stoffflussanalyse wie MFA, FBA und FVA im *CellNetAnalyzer* zur Verfügung.

5.2 Constrained Minimal Cut Sets

In diesem Abschnitt werden die constrained Minimal Cut Sets (cMCS) als eine alternative Methodik vorgestellt, um Interventionsstrategien zu identifizieren. In 5.2.1 wird zunächst die mathematische Beschreibung der constrained MCS als Erweiterung der MCS eingeführt und anhand eines Beispiels deren Anwendung demonstriert. Anschließend wird in 5.2.2 gezeigt, dass sich andere optimierungsbasierte und elementarmodenbasierte Ansätze als spezielle cMCS-Probleme darstellen und lösen lassen. Beispiele für die Formulierung komplexer Interventionsziele als cMCS-Probleme werden im Abschnitt 5.2.3 vorgestellt.

5.2.1 Definition von constrained Minimal Cut Sets

Wie zuvor gezeigt (siehe Kapitel 4.4.3), hat der bisherige Ansatz der MCS bzgl. der Anwendung für die Identifizierung von Interventionsstrategien einige Einschränkungen. Hier sei dies nochmals anhand des folgenden (größeren) Beispielnetzwerkes erläutert.

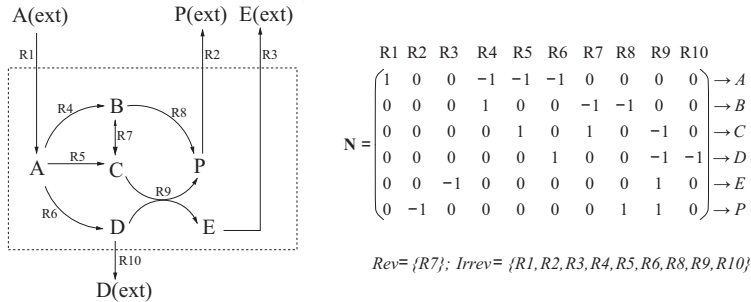


Abbildung 5.5: Beispielnetzwerk N_3

Angenommen, es sei das Ziel, für Netzwerk N_3 (Abbildung 5.5) Interventionen zu bestimmen, so dass der Metabolit P mit optimaler Ausbeute produziert wird. Für dieses Netzwerk existieren fünf EM, von denen zwei das Produkt P mit optimaler Ausbeute von 1 mol P pro mol A synthetisieren (Abbildung 5.6). Eine intuitive Interventionsstrategie, um das gewünschte Ziel zu erreichen, ist die Unterdrückung der Ausscheidung der Nebenprodukte D und E. Formal kann diese Aufgabe als MCS-Problem I_1 formuliert werden. Die Menge der Zielmoden ist in diesem Fall definiert durch $T = \{EM3, EM4, EM5\}$. EM werden hier wie zuvor über ihren Support, also als über die Menge der involvierten Reaktionen repräsentiert.

Es gibt sieben MCS, die dieses Problem lösen, darunter auch, als triviale Lösung, die Deletion der Substrataufnahmereaktion (Abbildung 5.7). Dieser MCS ist jedoch offensichtlich kein geeigneter Knockout, da er nicht nur die unerwünschten EM, sondern alle EM von N_3 und damit auch alle erwünschten Funktionalitäten deletiert. Diese Art der Beschränkung von geeigneten Knockout-Mengen tritt in vielen Anwendungen auf. Oftmals ist man an MCS interessiert, die zwar bestimmte Funktionalitäten deaktivieren, aber gleichzeitig andere Netzwerkfunktionalitäten erhalten.

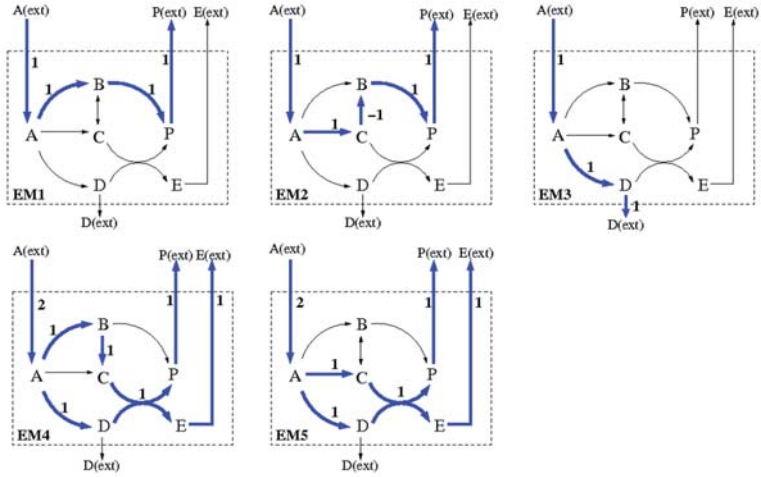


Abbildung 5.6: Elementarmoden des Netzwerkes N_3 . Die fettgedruckten blauen Linien kennzeichnen die beteiligten Reaktionen jedes EM. Zudem sind die relativen Flüsse entlang der beteiligten Reaktionen angegeben.

Um diese zusätzliche Bedingung zu berücksichtigen, wird nun das Konzept der constrained MCS eingeführt. Die grundlegende Idee dabei ist, MCS so zu berechnen, dass sowohl alle Zielmoden getroffen werden als auch eine bestimmte Anzahl gewünschter EM erhalten bleiben.

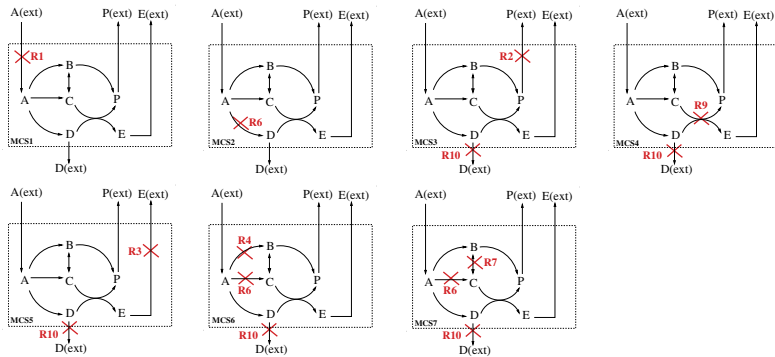


Abbildung 5.7: MCS zur Unterdrückung der Ausscheidung von D und E in N_3 ; sie entsprechen den Minimal Hitting Sets für die Targetmoden EM3, EM4 und EM5.

Die formale Definition eines MCS wird somit zunächst um eine Menge gewünschter Moden $\mathbf{D}$ erweitert. Für diese Elementarmoden, die eine Teilmenge aller EM $\mathbf{E}$ sind ($\mathbf{D} \subset \mathbf{E}$), wird hier auch die englische Bezeichnung Desired-EM verwendet. Man kann im Allgemeinen nicht davon ausgehen, dass alle EM in $\mathbf{D}$, die eine gewünschte Funktionalität widerspiegeln, durch einen MCS verschont bzw. nicht getroffen werden. Um eine gewünschte Funktionalität des Netzwerkes zu erhalten, genügt es meistens zu fordern, dass zumindest einige der gewünschten EM weiterhin zur Verfügung stehen. Durch die Spezifizierung einer unteren Schranke n , die eine minimale Anzahl von zu erhaltenen EM aus $\mathbf{D}$ angibt, können die Nebenbedingungen flexibel formuliert werden. In der Regel ist n also kleiner als die Anzahl der Moden, die in $\mathbf{D}$ enthalten sind, das heißt, $n \leq |\mathbf{D}|$. Für einen gegebenen MCS C , werden in $\mathbf{D}^C$ alle EM $D \in \mathbf{D}$ gesammelt, die nicht durch C betroffen sind:

$$\mathbf{D}^C = \{D \in \mathbf{D} \mid C \cap D = \emptyset\}. \quad (5.11)$$

Ein constrained MCS erfüllt Gleichung 4.11 bzgl. der Targetmoden $\mathbf{T}$ sowie die Minimalitätsforderung und die Nebenbedingung:

$$|\mathbf{D}^C| \geq n. \quad (5.12)$$

Interventionsprobleme $I = (\mathbf{T}, \mathbf{D}, n)$ können somit definiert werden mittels:

1. Einer Menge von Targetmoden $\mathbf{T}$,
2. einer Menge gewünschter Moden $\mathbf{D}$,
3. einer Mindestanzahl n von gewünschten Moden, die erhalten bleiben sollen.

Die ursprüngliche Definition von MCS ist in dieser Formulierung enthalten, indem für $\mathbf{D} = \emptyset$ und $n = 0$ gewählt wird. Für bestimmte Anwendungen kann es nützlich sein, mehrere Mengen $\mathbf{D}_1, \dots, \mathbf{D}_g$ mit zugehörigen $n_1, \dots, n_g$ festzulegen und somit zu fordern, dass aus jeder Teilmenge $\mathbf{D}_i$ eine bestimmte Anzahl Moden erhalten bleibt. Auf diese Weise kann eine logische UND-Verknüpfung dieser Bedingungen realisiert werden. Dementprechend ist ein Interventionsproblem definiert durch:

$$I = (\mathbf{T}, \mathbf{D}_1, n_1, \mathbf{D}_2, n_2, \dots, \mathbf{D}_g, n_g). \quad (5.13)$$

Ein gültiger cMCS C muss die Bedingung 4.11 sowie die Minimalitätsforderung erfüllen und zudem muss für jedes Paar $(\mathbf{D}_i, n_i)$ gelten:

$$|\mathbf{D}_i^C| \geq n_i, \quad i = 1, \dots, g. \quad (5.14)$$

Sinnvoll aufgestellte Interventionsprobleme erfüllen hierbei, dass ein EM nicht gleichzeitig sowohl Target- als auch Desired-EM ist. Also gilt:

$$\mathbf{D}_i \cap \mathbf{T} = \emptyset. \quad (5.15)$$

Es ist durchaus möglich, dass einige EM aus **E** keiner der Mengen **T** oder **D** zugeteilt sind. Die Vereinigung von **T** und **D** muss somit nicht zwingend die vollständige Menge **E** abdecken. Die nicht zugewiesenen EM werden als „neutral“ betrachtet. Sie enthalten keine ungewünschte Charakteristik; müssen mithin also nicht deletiert werden, sind gleichzeitig aber auch nicht von besonderem Interesse, so dass es keines dringenden Erhalts bedarf. Ein typisches Beispiel sind EM, die eine annähernd optimale Produktausbeute bei niedriger Biomasseausbeute ermöglichen. Solche Flusslagen werden vermutlich nicht von der Zelle angestrebt und müssen daher nicht deletiert werden, zeigen aber auch oftmals nicht die beste Performance, als dass man sie unbedingt erhalten sollte.

Mit diesen Erweiterungen soll erneut das Ziel der optimalen Produktion von P im Beispielnetzwerk N_3 (Abbildung 5.5) verfolgt werden. Das ursprüngliche Interventionsproblem I_1 (Tabelle 5.2) wird um die Forderung erweitert, dass eine optimale Produktion von P gewährleistet wird, während die Synthese von E und D unterbunden wird. Für dieses Interventionsproblem I_2 sind die Target-Moden identisch zu Problem I_1 . Jedoch wird zusätzlich verlangt, dass mindestens einer ($n = 1$) der beiden Moden erhalten bleibt, die die Herstellung von P mit maximaler Ausbeute realisieren. Die Menge der Desired-EM ist damit gegeben durch $\mathbf{D} = \{\text{EM1}, \text{EM2}\}$. Wie zu erwarten war, sind die resultierenden vier zulässigen cMCS eine Teilmenge der MCS vom ersten Szenario, da die gleiche Menge der Target-Moden getroffen werden muss (Tabelle 5.2). Wenn die gleichen Target-Moden eliminiert werden sollen, sind cMCS immer Teilmengen der zugehörigen MCS.

Tabelle 5.2: Definition und Lösungen verschiedener cMCS-Probleme für N_3 . Die MCS der Probleme I_1 , I_2 und I_3 entsprechen denen aus Abb. 5.7.

Problem	D	T	n	MCS
I_1	$\emptyset$	$\{\text{EM3}, \text{EM4}, \text{EM5}\}$	0	MCS1-MCS7
I_2	$\{\text{EM1}, \text{EM2}\}$	$\{\text{EM3}, \text{EM4}, \text{EM5}\}$	1	MCS2, MCS4, MCS5, MCS7
I_3	$\{\text{EM1}, \text{EM2}\}$	$\{\text{EM3}, \text{EM4}, \text{EM5}\}$	2	MCS2, MCS4, MCS5
I_4	$\mathbf{D}_1 = \{\text{EM1}, \text{EM2}\}$ $\mathbf{D}_2 = \{\text{EM3}\}$	$\{\text{EM4}, \text{EM5}\}$	$n_1 = 1$ $n_2 = 1$	$\{\text{R3}\}, \{\text{R9}\},$ $\{\text{R5}, \text{R7}\}$

Ein drittes Interventionsproblem, I_3 , wird erhalten, indem im Gegensatz zu I_2 verlangt wird, dass alle EM mit maximaler Ausbeute von P funktional bleiben ($n = 2$). Die drei zulässigen cMCS sind wiederum eine Teilmenge der Lösungen von I_2 (und I_1). Einer der gültigen cMCS, das dieses Problem löst, benötigt lediglich die Deletion von R6 (MCS2).

Ein Beispiel zur Definition eines Interventionsproblems mit mehr als einer Menge von Desired-Moden ist mit I_4 gegeben (Tabelle 5.2). In diesem wird verlangt, dass die Produktion von Metabolit E unterdrückt werden soll, während die Synthese der Metabolite P und D weiterhin mit der maximalen Ausbeute möglich sein soll. Die Menge der Target-Moden umfasst nun die Elementarmoden EM4 und EM5. Die Menge der Desired-Moden besteht aus zwei Teilmengen $\mathbf{D}_1$ und $\mathbf{D}_2$, wobei $\mathbf{D}_1 = \{\text{EM1}, \text{EM2}\}$ die Moden zur opti-

malen Produktion von P umfasst und $\mathbf{D}_2 = \{\text{EM3}\}$ den Modus zur optimalen Produktion von D enthält. Das Interventionsziel kann also nur erreicht werden, wenn von jeder dieser Teilmengen mindestens ein EM erhalten bleibt. Es können drei zulässige cMCS bestimmt werden, die keine Teilmenge der Lösungen von I_1 sind, da sich auch die Mengen der Target-Moden der Interventionsprobleme unterscheiden (Tabelle 5.2).

Die Randbedingungen (Gleichung 5.14) bei der Berechnung der cMCS mit dem Berge-Algorithmus können auf zwei verschiedene Weisen berücksichtigt werden. Ein einfacher Ansatz ist, alle MCS wie üblich zu berechnen und in einem nachgelagerten Schritt alle MCS zu verwerfen, welche die Bedingungen für ein gegebenes Paar $(\mathbf{D}_i, n_i)$ verletzen. Es hat sich jedoch gezeigt, dass es in typischen Anwendungen vorteilhaft ist, wenn nach jeder Iteration des Berge-Algorithmus' geprüft wird, ob die Bedingungen für alle Paare $(\mathbf{D}_i, n_i)$ erfüllt sind. Dies gilt insbesondere, wenn die Schwellenwerte n_i hoch sind und bzw. oder die Kardinalitäten der Mengen der Desired-EM $\mathbf{D}_i$ klein sind. Auf diese Weise können viele vorläufige cMCS bereits während früher Iterationen verworfen werden, so dass diese nicht zu MCS höherer Ordnung kombiniert werden. Der Pseudo-Code des angepassten Berge-Algorithmus zur Berechnung der cMCS ist in Abbildung 5.8 gegeben.

```

// Argumente:
// 1) Menge T der Target-EM (jeder EM ist repräsentiert als Teilmenge der Reaktionen),
// 2) Mengen Di der Desired-EM,
// 3)  $n_i$  als die Anzahl der mindestens zu erhaltenen EM der zugehörigen Di,
// 4) maximale Anzahl der erlaubten Knockouts maxMCSsize (kann unbeschränkt sein).
// Rückgabe:
// Menge M der zulässigen cMCS

M = { { } }; // als leere Menge initialisieren
für jeden EM T aus T
    für jeden vorläufigen cut set M aus M
        wenn  $T \cap M = \emptyset$  // die vorläufigen Cut Sets treffen T nicht
            M aus M entfernen;
            wenn  $(|M| < \text{maxMCSsize})$  //Anzahl Knockouts beachten
                für jede Reaktion k in T
                    füge  $(M \cup k)$  zu M hinzu; //kreiere neue cMCS-Kandidaten
                end
            end
        end
    end
    Test auf Minimalität aller neuen M aus M, verwerfe nicht minimale M;
    Test der Nebenbedingungen  $(\mathbf{D}_i, n_i)$  aller neuen M aus M, verwerfe unzulässige M;
end

// Die Nebenbedingungen bzgl.  $(\mathbf{D}_i, n_i)$  können alternativ zum Test nach jeder Iteration
// hier in einem Nachverarbeitungsschritt überprüft werden.

Ergebnis: M enthält alle zulässigen cMCS.
```

Abbildung 5.8: Pseudo-Code des adaptierten Berge-Algorithmus zur Berechnung von constrained MCS.

Grundsätzlich lassen sich cMCS Probleme auch als binäre Integer-Optimierungsprobleme (BIP) lösen (Jungreuthmayer und Zanghellini, 2012). Durch die Verwendung von Binärbäumen und diversen Vorverarbeitungsschritten kann die Laufzeit beider Algorithmen (Berge und BIP) deutlich verbessert werden, wobei der adaptierte Berge-Algorithmus stets die besseren Laufzeitergebnisse liefert (Jungreuthmayer et al., 2013).

5.2.2 Reformulierung anderer Ansätze

Die große Flexibilität des cMCS-Ansatzes erlaubt es, viele Methoden als spezielle cMCS-Probleme zu formulieren. Im Folgenden wird dies für drei populäre Beispiele erläutert.

Minimale Metabolische Funktionalität (MMF)

Eine konkrete MMF-Lösung lässt sich retrospektiv sehr einfach durch die Lösung eines zugehörigen cMCS-Problems wiederfinden. Die Menge der Desired-EM wird so definiert, dass sie genau der Menge der nicht deletierten EM (mit der MMF-Methodik) entspricht. Alle verbleibenden EM werden der Targetmenge **T** zugeordnet, so dass jeder EM entweder Element der Menge **D** oder **T** ist. Schließlich wird gefordert, dass alle EM der Menge **D** funktional bleiben sollen, also $n = |\mathbf{D}|$. Die ursprüngliche MMF-Lösung ist dann eine Teilmenge aller cMCS-Lösungen. So kann gezeigt werden, dass eine per MMF identifizierte Siebenfach-Knockout-Strategie zur Produktion von Ethanol mit *E. coli* (Trinh et al., 2008) genau eine Lösung von 252 möglichen verschiedenen Siebenfach-Knockout-Strategien ist. Insgesamt existieren für dieses Beispiel 1048 cMCS-Lösungen, darunter auch 60 Lösungen, die nur 6 Reaktionsdeletionen benötigen. Dieses Beispiel demonstriert einen zentralen Vorteil der cMCS-Methode. Es wird nicht nur eine optimale Knockout-Strategie berechnet, sondern es werden alle äquivalenten Lösungen ermittelt.

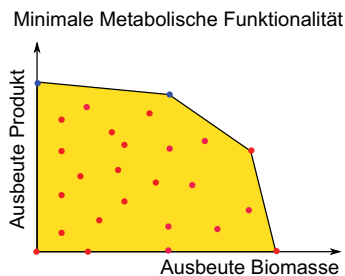


Abbildung 5.9: MMF als Spezialfall eines cMCS-Problems. Typischerweise müssen für MMF-Probleme wenige Desired-EM (blaue Punkte) erhalten bleiben. Alle verbleibenden EM sind Target-EM (rote Punkte) und müssen deletiert werden.

In der ursprünglichen MMF-Formulierung stehen die konkret zu erhaltenen EM nicht explizit am Beginn der Methode fest. Somit ist es *a priori* nicht möglich, ein zugehöriges eindeutiges cMCS-Problem zu definieren. Dennoch muss auch für die MMF-Methode vor Beginn festgelegt werden, welche Funktionalitäten bzw. Eigenschaften erhalten bleiben sollen. Dies lässt sich leicht in die Formulierung eines cMCS-Problems übertragen. So

werden in der Menge der Desired-EM $\mathbf{D}$ einige wenige EM mit gewünschten Eigenschaften definiert. Alle verbleibenden EM werden der Menge der Target-Moden $\mathbf{T}$ zugeteilt. Je nach konkreter Anwendung wird zudem gefordert, dass mindestens ein und höchstens alle EM aus $\mathbf{D}$ erhalten bleiben: $1 \leq n \leq |\mathbf{D}|$. Mit Hilfe dieser Formulierung (Abbildung 5.9 und Tabelle 5.3) können somit verallgemeinerte MMF-Probleme als analoge cMCS-Probleme definiert und gelöst werden.

OptKnock und RobustKnock

Die Formulierung der OptKnock-Methodik als cMCS-Problem ist weniger intuitiv, aber dennoch möglich. So werden aus der Menge aller EM, deren Biomasseausbeuten über einem Grenzwert $t_{X/S}$ liegen, alle EM der Menge $\mathbf{D}$ zugeordnet, die eine maximale Produktausbeute ermöglichen. Wenn mehrere EM das optimale Verhalten zeigen, muss mindestens einer dieser EM erhalten werden. Der Menge $\mathbf{T}$ werden alle EM zugeordnet, deren Biomasse-Ausbeute echt größer als die Biomasseausbeute der EM aus $\mathbf{D}$ ist (Tabelle 5.3). Auf diese Weise wird sichergestellt, dass die maximale Biomasseausbeute mit einem EM aus der Menge der Desired-EM erreicht wird. Zusätzlich wird garantiert, dass einer der EM mit maximaler Biomasseausbeute auch die maximale Produktausbeute gewährleistet (Abbildung 5.10).

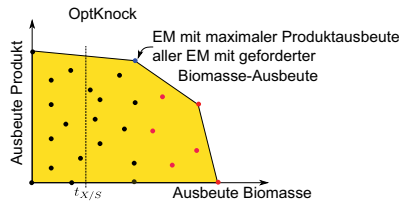


Abbildung 5.10: OptKnock als Spezialfall eines cMCS-Problems.

Für die Formulierung des RobustKnock-Ansatzes muss lediglich die Menge der Target-EM noch um die EM erweitert werden, deren Biomasseausbeute der Biomasseausbeute der Desired-EM mit Gleichheit entsprechen, deren Produktausbeute aber kleiner als der Desired-EM ist (Tabelle 5.3). Diese Bedingung garantiert, dass es keinen EM geben kann, der zwar eine optimale Biomasseausbeute, aber nicht gleichzeitig auch die optimale Produktausbeute ermöglicht (Abbildung 5.11).

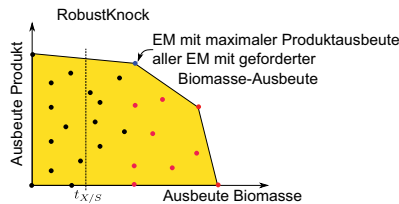


Abbildung 5.11: RobustKnock als Spezialfall eines cMCS-Problems.

Tabelle 5.3: Formulierung verschiedener Methoden als cMCS-Probleme. X Biomasse, S Substrat, $t_{X/S}$ Mindest-Biomasseausbeute, P Produkt, Y_E Ausbeute des EM E , $\mathbf{R} = \{E \in \mathbf{E} | Y_E^{X/S} \geq t_{X/S}\}$.

	Desired-EM $\mathbf{D}$	Target-EM $\mathbf{T}$	n
MMF	Spezifische Menge EM mit gewünschten Eigenschaften	$\mathbf{E} \setminus \mathbf{D}$	$[1, \mathbf{D}]$
OptKnock	$\left\{ E \in \mathbf{R} \mid Y_E^{P/S} = \max_{H \in \mathbf{R}} (Y_H^{P/S}) \right\}$	$\left\{ E \in \mathbf{E} \mid Y_E^{X/S} > \max_{D \in \mathbf{D}} (Y_D^{X/S}) \right\}$	1
RobustKnock	$\left\{ E \in \mathbf{R} \mid Y_E^{P/S} = \max_{H \in \mathbf{R}} (Y_H^{P/S}) \right\}$	$\left\{ E \in \mathbf{E} \mid Y_E^{X/S} \geq \max_{D \in \mathbf{D}} (Y_D^{X/S}) \setminus \mathbf{D} \right\}$	1

Im Gegensatz zur ursprünglichen Definition von OptKnock und RobustKnock ist durch diese Formulierung die Anzahl der maximal erlaubten Knockouts nicht beschränkt. Solange die Anzahl der erlaubten Knockouts größer ist als die Kardinalität der kleinsten cMCS-Lösung, sind die Lösungen von OptKnock und RobustKnock Teilmengen der Lösungen der jeweils zugehörigen cMCS-Probleme. Da über den Grenzwert $t_{X/S}$ (für die geforderte minimale Biomasseausbeute) die Menge der Desired- und Target-EM bestimmt wird, ist auch die Kardinalität der kleinsten Lösung von diesem Grenzwert abhängig.

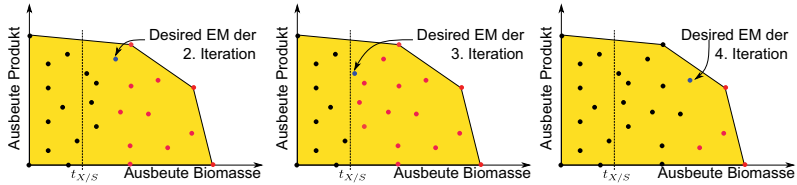


Abbildung 5.12: Iteratives Vorgehen bei Beschränkung der Anzahl der Knockouts. In jeder Iteration wird (wenn die vorhergehende Iteration keine Lösung lieferte) der nachfolgend ausbeuteoptimale EM mit ausreichender Biomasseausbeute als Desired-EM gewählt und die Menge der Target-EM entsprechend gewählt.

Ist die Anzahl der erlaubten Knockouts kleiner als notwendig, um die maximale Ausbeute zu ermöglichen, kann der Wert der realisierbaren Ausbeute durch ein iteratives Vorgehen bestimmt werden. Dafür wird zunächst die maximale cMCS-Größe auf die Anzahl der erlaubten Knockouts beschränkt. Wenn keine Lösungen für dieses Problem existieren, werden die aktuellen EM aus $\mathbf{D}$ entfernt und durch den EM ersetzt, der die nächstkleinere Produktausbeute bei gleichzeitiger Mindestbiomasseausbeute ermöglicht. Anschließend wird die Menge $\mathbf{T}$ entsprechend angepasst und das cMCS-Problem erneut gelöst. Dieses Vorgehen wird solange wiederholt, bis eine Lösung bestimmt werden kann (Abb. 5.12).

5.2.3 Verallgemeinerung für komplexe Interventionen

Neben der exakten Reformulierung von bekannten Algorithmen können mit dem cMCS-Ansatz beliebige komplexe Interventionsprobleme formuliert werden. Es sei daran erinnert, dass die zu Grunde liegende Idee fast aller Methoden ist, den möglichen Bereich des phänotypischen Verhaltens auf den Gewünschten zu reduzieren. Dieser gewünschte Bereich kann sehr bequem mittels der Desired-EM beschrieben werden. Indem bestimmte EM, die außerhalb dieses Bereiches liegen, als Target-Moden definiert werden, kann zugleich garantiert werden, dass keine Flussverteilungen mit Ausbeuteverhältnissen verbleiben, die nicht diesem Bereich entsprechen.

Mit Hilfe von zwei Grenzwerten können dabei verschiedene Bereiche der Phänotypenebene in die Bereiche der Desired-EM, Target-EM oder unklassifizierter EM unterteilt werden. Vier verschiedene Unterteilungen sind dabei von besonderem Interesse und werden hier kurz erläutert. Für die vier diskutierten Fälle ist die Existenz und die Anzahl von Lösungen, neben den gewählten Grenzwerten, abhängig von der Anzahl der zu erhaltenen Desired-EM. Im Folgenden werden die Auswirkungen der jeweiligen cMCS-Formulierungen abstrakt diskutiert. Konkrete Fallbeispiele werden im nächsten Kapitel vorgestellt.

$$\mathbf{1. \text{ Fall: } D = \{E \in \mathcal{E} | Y_E^{X/S} \geq t_{X/S} \wedge Y_E^{P/S} \geq t_{P/S}\}, \quad T = \{E \in \mathcal{E} | Y_E^{X/S} \geq t_{X/S} \wedge Y_E^{P/S} < t_{P/S}\}}$$

Im ersten Fall wird für die Desired-EM gefordert, dass sowohl deren Biomasse- als auch Produktausbeute über spezifischen Grenzwerten $t_{X/S}$ und $t_{P/S}$ liegt. In die Menge der Target-EM werden alle EM selektiert, die zwar eine hohe Biomasseausbeute, aber nur eine ungenügende Produktausbeute erreichen. Somit werden alle EM, deren Biomasseausbeute kleiner als $t_{X/S}$ ist, weder den Desired- noch den Target-EM zugeteilt (Abb. 5.13).

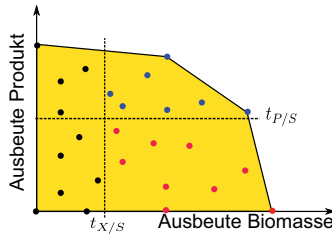


Abbildung 5.13: Unterteilung der EM nach Fall 1: Desired-EM (blaue Punkte), Target-EM (rote Punkte), unklassifizierte EM (schwarze EM).

Ein Argument für diese Art der EM-Einteilung ist die Annahme, dass der Organismus im Laufe der Zeit seine Flussverteilung so organisiert, dass hauptsächlich EM mit hoher (optimaler) Biomasseausbeute verwendet werden. Die EM mit niedriger Biomasseausbeute müssten daher nicht zwingend deletiert werden.

Die Phänotypenebenen der möglichen Mutanten können allerdings ähnliche Schwächen aufweisen, wie die von mittels OptKnock/RobustKnock identifizierten Mutanten (Abbildung 5.14). So kann für suboptimale Wachstumsbedingungen keine untere Schranke für

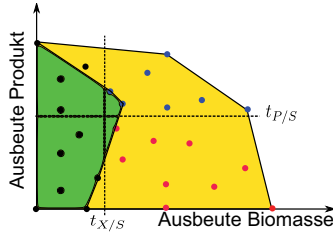


Abbildung 5.14: Eine mögliche Phänotypenebene (grüner Bereich) einer Mutante nach Knockout-Strategie aus Fall 1.

die Produktausbeute garantiert werden. Zudem kann die Produktkopplung für suboptimales Wachstum komplett verschwinden. Allerdings stellt die in Abbildung 5.14 dargestellte Mutante den „worst-case“ dar. Blieben EM aus **D** mit einer höheren Biomasseausbeute erhalten, so wäre für einen großen Bereich suboptimalen Wachstums dennoch eine Produktausscheidung garantiert. Ebenso wäre die Definition eines weiteren Grenzwertes $t_{X/S}^D > t_{X/S}$ denkbar. Dieser würde dann für die Selektion der EM für **D** verwendet, so dass ein zu steiler Abfall der garantierten Kopplung für suboptimales Wachstum vermieden werden könnte.

Obwohl alle EM des Bereiches mit $Y^{X/S} \geq t_{X/S} \wedge Y^{P/S} < t_{P/S}$ deletiert wurden, ist es möglich, dass die Flusslage der Mutante sich in diesen Bereich verschiebt. Daher sollte geprüft werden, ob mit alternativen Problemformulierungen Interventionsstrategien identifiziert werden können, die dieses Verhalten ausschließen. Der Vorteil dieser Formulierung ist die relativ kleine Menge der Target-EM, so dass typischerweise Interventionsstrategien mit wenigen Knockouts gefunden werden. Es sei nochmals darauf hingewiesen, dass Knockouts hier im Sinne von Reaktionsdeletionen und nicht als Gen-Knockouts identifiziert werden.

2. Fall: $\mathbf{D} = \{E \in \mathbf{E} | Y_E^{P/S} \geq t_{P/S}\}$, $\mathbf{T} = \{E \in \mathbf{E} | Y_E^{P/S} < t_{P/S}\}$

In diesem Fall wird für die Desired-EM nur gefordert, dass sie eine bestimmte Mindestausbeute garantieren. Entsprechend werden alle EM, deren Produktausbeute kleiner als ein bestimmter Grenzwert $t_{P/S}$ ist, als Target-EM klassifiziert. Somit sind in diesem Fall alle EM entweder in **D** oder in **T** enthalten (Abbildung 5.15).

Die Phänotypenebene der besten Mutante ist in diesem Fall durch die konvexe Hülle der Desired-EM begrenzt (Abbildung 5.16). Im schlechtesten Fall könnten nur EM ohne Biomassebildung erhalten bleiben, so dass diese cMCS-Formulierung ohne weitere Bedingungen nicht zwingend zu Interventionsstrategien führt, die das gewünschte Verhalten ermöglichen. Wird zusätzlich gefordert, dass mehr EM erhalten bleiben müssen als in **D** ohne Biomassebildung enthalten sind ($n > |\{EM \in \mathbf{D} \wedge Y^{X/S} = 0\}|$), kann zumindest der „worst-case“ ausgeschlossen werden. Der Vorteil dieser Formulierung ist, dass die zugehörigen Interventionsstrategien eine Mindestproduktausbeute unabhängig vom Wachstumsverhalten garantieren können. Wird n nicht zu groß gewählt $n \ll |\mathbf{D}|$, so werden auch mit dieser Formulierung meist einige Interventionsstrategien mit wenigen Knockouts identifiziert werden können.

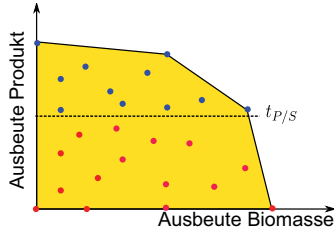


Abbildung 5.15: Unterteilung der EM nach Fall 2: Desired-EM (blaue Punkte), Target-EM (rote Punkte).

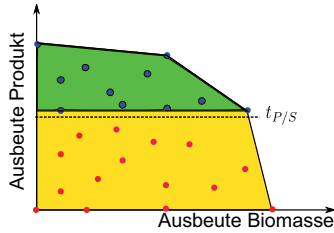


Abbildung 5.16: Eine mögliche Phänotypebene (grüner Bereich) einer Mutante nach Knockout-Strategie aus Fall 2.

3. Fall: $\mathbf{D} = \{E \in \mathbf{E} | Y_E^{X/S} \geq t_{X/S} \wedge Y_E^{P/S} \geq t_{P/S}\}$, $\mathbf{T} = \{E \in \mathbf{E} | Y_E^{X/S} < t_{X/S}\} \cup \{E \in \mathbf{E} | Y_E^{P/S} < t_{P/S}\}$

Eine dritte plausible Einteilung der EM kann vorgenommen werden, indem die Menge der Desired-EM wie in Fall 1 aus den EM mit hoher Biomasse- und Produktausbeute besteht. Als Target-EM werden hier alle verbleibenden EM eingeteilt, so dass alle EM entweder in $\mathbf{D}$ oder in $\mathbf{T}$ enthalten sind (Abbildung 5.17).

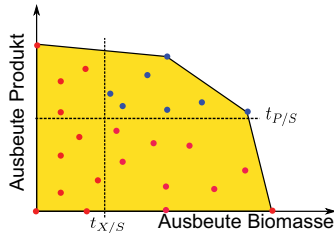


Abbildung 5.17: Unterteilung der EM nach Fall 3: Desired-EM (blaue Punkte), Target-EM (rote Punkte).

Ein Vorteil dieser Klassifizierung ist, dass die resultierenden Phänotypebenen aller möglichen Mutanten auf den gewünschten Bereich begrenzt sind (Abbildung 5.18). Falls Lösungen identifiziert werden können, ist somit garantiert, dass die Flusslage der Mutante sich in dem gewünschten Bereich etabliert. Als Nachteil kann angeführt werden, dass

auch die EM mit optimaler Produktausbeute deletiert werden müssen. Zudem können durch die relativ große Menge an Target EM viele Knockouts notwendig sein.

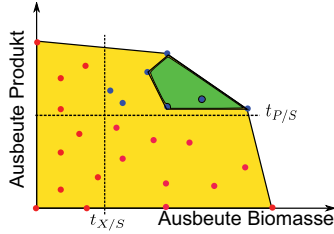


Abbildung 5.18: Eine mögliche Phänotypebene (grüner Bereich) einer Mutante nach Knockout-Strategie aus Fall 3.

4. Fall: $\mathbf{D} = \{E \in \mathbf{E} | Y_E^{X/S} \geq t_{X/S} \wedge Y_E^{P/S} \geq t_{P/S}\}$, $\mathbf{T} = \{E \in \mathbf{E} | Y_E^{P/S} < t_{P/S}\}$

Analog zum ersten Fall wird für die Desired-EM gefordert, dass sowohl deren Biomasse- als auch Produktausbeute über spezifischen Grenzwerten $t_{X/S}$ und $t_{P/S}$ liegt. In die Menge der Target-EM werden hier aber alle EM selektiert, die eine nur ungenügende Produktausbeute erreichen. Somit ist auch die Menge der EM, die weder $\mathbf{D}$ noch $\mathbf{T}$ zugeteilt sind, kleiner als im Fall 1 und beschränkt sich auf die EM, deren Biomasseausbeute kleiner als $t_{X/S}$ ist und deren Produktausbeute größer als $t_{P/S}$ ist (Abbildung 5.19).

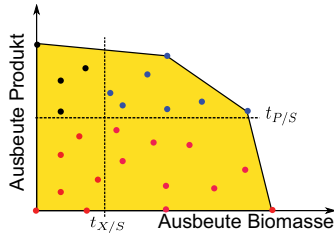


Abbildung 5.19: Unterteilung der EM nach Fall 4: Desired-EM (blaue Punkte), Target-EM (rote Punkte), unklassifizierte EM (schwarze EM).

Die EM mit hoher Produkt-, aber niedriger Biomasseausbeute werden auch hier analog zur Argumentation in Fall 1 nicht deletiert, da angenommen werden kann, dass diese nicht primär von der Mutante verwendet werden. Vielmehr ist es so, dass es durchaus erwünscht ist, dass EM weiterhin funktional bleiben, wenn diese optimale Produktausbeuten ermöglichen. Die realisierbaren Flusslagen der Mutante garantieren somit unabhängig vom Wachstumsverhalten die geforderte Mindestproduktausbeute, während minimale Biomasseausbeuten möglich sind (Abbildung 5.20).

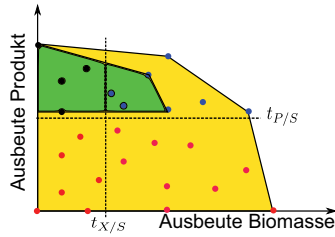


Abbildung 5.20: Eine mögliche Phänotypebene (grüner Bereich) einer Mutante nach Knockout-Strategie aus Fall 4.

Ausbeute versus Produktivität

Die cMCS-Methodik beinhaltet keinen expliziten Parameter, um den Schwerpunkt der gewünschten Strategien auf die Ausbeute- oder Produktivitätsoptimalität zu legen. Dennoch ist es möglich, die Probleme so zu formulieren, dass den beiden Kriterien ein unterschiedlich starkes Gewicht zugeteilt wird.

Liegen zum Beispiel Informationen über die Kapazität einzelner EM vor, so kann gefordert werden, dass ausschließlich solche mit einer bestimmten Mindestkapazität im Netzwerk erhalten bleiben. Ohne diese Informationen kann analog zur Argumentation der CASOP-Heuristik gefordert werden, dass eine möglichst große Anzahl an erwünschten EM erhalten bleibt. Zudem können weniger stringente Anforderungen an die angestrebte Ausbeute gestellt werden, so dass mehr suboptimale Routen erhalten bleiben. Die dadurch verbleibende hohe Flexibilität des Netzwerkes dient wiederum als Approximation der Kapazität bzw. des potentiellen Durchsatzes.

Strategien mit Fokus auf der Optimalität der Ausbeute können bestimmt werden, indem gefordert wird, dass bis auf wenige, ausbeuteoptimale EM alle Routen im Netzwerk deletiert werden. Dieser Ansatz ist analog zur Idee der minimalen metabolischen Funktionalität.

Implementierung

Die bereits vorhandene MCS-Funktionalität des *CellNetAnalyzer* wurde um die cMCS-Methodik erweitert. Verschiedene Auswahlkriterien (z.B. Verwendung bestimmter Reaktionen oder Metaboliten, minimale bzw. maximale Ausbeuten etc.) erlauben es dem Benutzer, per GUI bequem eine Menge von Target-EM und bis zu zwei Teilmengen von Desired-EM mit zugehörigen Parametern n_1 und n_2 zu definieren. In der Kommandozeilenversion dieser Funktion können beliebig viele Paare $(\mathbf{D}_i, n_i)$ definiert werden. Zudem kann die Anzahl der maximal erlaubten Knockouts (die Kardinalität der cMCS) durch den Benutzer spezifiziert werden. Die berechneten cMCS können in der grafischen Benutzeroberfläche visualisiert und weiter analysiert werden, z. B. auf essentielle Reaktionsdeletionen, die in allen cMCS enthalten sein müssen.

5.3 EFMSampler

Der EFMSampler, der in diesem Abschnitt vorgestellt wird, ist keine unmittelbare Methode für die Identifizierung von Interventionsstrategien. Vielmehr handelt es sich um einen Ansatz zur Berechnung von Teilmengen von Elementarmoden in genomskaligen Netzwerken. Da es für einige Anwendungen ausreichend ist, eine repräsentative Teilmenge der Elementarmoden zu berechnen, eröffnet der EFMSampler damit die Möglichkeit, elementarmodenbasierte Verfahren (wie z.B. CASOP) auch für solche großskaligen Netzwerke anwenden zu können.

5.3.1 Elementarmodenberechnung für genomskalige Netzwerke

Die Enumeration aller EM ist beschränkt durch die Größe der betrachteten Netzwerke. Für genomskalige Netzwerke ist es meistens auf Grund der kombinatorischen Komplexität (die Anzahl der EM wächst exponentiell mit der Anzahl der Freiheitsgrade eines Netzwerkes) nicht möglich, die vollständige Menge aller EM zu berechnen (Klamt und Stelling, 2002; Acuña et al., 2010). Ein erster Ansatz, um ausgewählte EM in genomskaligen Netzwerken zu berechnen, wurde von (de Figueiredo et al., 2009) vorgeschlagen. Dabei werden mittels einer MILP-basierten Methode die kürzesten EM eines Netzwerkes berechnet. Mit diesem Ansatz können jedoch nur einige hundert EM in vertretbarer Zeit bestimmt werden (de Figueiredo et al., 2009; Gebauer et al., 2012).

Eine andere mögliche Herangehensweise besteht darin, eine repräsentative Stichprobe von EM zu ermitteln. Ähnliche Ideen, die ebenfalls eine Beschreibung des Lösungsraumes durch Stichprobenansätze verfolgen, wurden durch (Wiback et al., 2004), (Almaas et al., 2004) und (Machado et al., 2012b) beschrieben. Der EFMEvolver verfolgt einen alternativen Ansatz. Diese Methode berechnet mittels linearer Optimierung und eines genetischen Algorithmus EM in großskaligen Netzwerken (Kaleta et al., 2009). Diese Methode wurde bereits angewendet, um Zyklen in einem genomskaligen Stoffwechselnetzwerk des Menschen zu identifizieren (Gebauer et al., 2012). Im Folgenden wird der EFM-Sampler vorgestellt, eine Weiterentwicklung dieser Methode. Dieser verwendet an Stelle des genetischen Algorithmus eine iterative Heuristik zur Berechnung einer repräsentativen Teilmenge von EM.

5.3.2 EFMSampler zur Berechnung von Elementarmodenteilmengen

Wie bereits in Kapitel 4.3.2 eingeführt, entsprechen die Extremalstrahlen des zulässigen Flusskegels (der durch die Annahme des Fließgleichgewichts und den Irreversibilitätsnebenbedingungen definiert ist) einer Teilmenge der EM. Wenn zusätzlich alle reversiblen Reaktionen in je zwei irreversible Reaktionen aufgespalten werden, so existieren keine EM im Inneren des Kegels und die Menge der Extremalstrahlen entspricht der Menge der EM. Diese Eigenschaft erlaubt es, ein lineares Optimierungsproblem zu formulieren, bei dem jede Ecke der beschränkten Lösungsmenge genau einem EM entspricht. Das kann dadurch erreicht werden, dass der Fluss durch eine beliebig wählbare Reaktion t (z.B. die Biomassessynthese oder die Produktausscheidung) auf einen festen Wert fixiert wird.

Mittels dieser zusätzlichen Nebenbedingung entsprechen die Ecken des zulässigen und beschränkten Lösungsraumes dann den EM des Netzwerkes, bei dem Reaktion t aktiv ist.

Das lineare Optimierungsproblem (LP) ist wie folgt definiert:

$$\begin{aligned} \mathbf{N} \mathbf{r} &= \mathbf{0} \\ \mathbf{r} &\geq \mathbf{0} \\ r_t &= 1. \end{aligned} \tag{5.16}$$

Eine Zielfunktion muss nicht definiert werden, da das LP nur zur Generierung der EM und nicht zur Maximierung einer speziellen Reaktionsrate verwendet werden soll. Es genügt demnach, nur die erste Phase des LP-Problems zu lösen, bei dem lediglich ein zulässiger Startwert identifiziert wird. Durch die Verwendung des Simplex-Algorithmus (Dantzig, 1998) wird gewährleistet, dass jeder gefundene zulässige Punkt stets eine Ecke des Lösungsraumes ist und somit einem EM entspricht. Da das LP-Problem gelöst ist, sobald ein zulässiger Punkt gefunden wird, kann genau ein EM berechnet werden, der die Reaktion t verwendet. Um weitere EM zu berechnen, werden mittels eines iterativen Vorgehens zufällig ausgewählte Reaktionen auf Null fixiert (deletiert) und anschließend das LP-Problem erneut gelöst. Im Folgenden wird auf dieses Vorgehen detailliert eingegangen.

Heuristik zur Bestimmung der blockierten Reaktionen

Nachdem der EFMSampler einen ersten EM gefunden hat, wird der Fluss durch zufällig ausgewählte Reaktionen $k_1, k_2, \dots, k_l$ auf Null fixiert (blockiert) und das lineare Optimierungsproblem erneut gelöst. Aus $l \geq n - \text{rang}(\mathbf{N}) - 1$ vielen Reaktionen können dafür Kandidaten gewählt werden, da ein EM höchstens $\text{rang}(\mathbf{N}) + 1$ nicht Null Einträge enthält. Indem der Fluss durch diese Reaktionen auf Null beschränkt wird, bleibt die Eigenschaft erhalten, dass die Kanten gerade den EM entsprechen. Eine zulässige Lösung des modifizierten linearen Optimierungsproblems liefert somit auch einen neuen EM.

Zu Beginn der Heuristik ist die Menge der blockierten Reaktionen (Q_0) leer, $Q_0 = \emptyset$. Durch Lösen des LP wird ein erster EM berechnet. Für die nächste Iteration ($h + 1$) wird eine Reaktion i zufällig ausgewählt und zur Menge der blockierten Reaktionen hinzugefügt $Q_{h+1} = Q_h \cup i$. Anschließend wird das lineare Optimierungsproblem 5.16 mit den zusätzlichen Nebenbedingungen

$$\forall i \in Q_h : r_i = 0 \tag{5.17}$$

gelöst, um weitere EM zu berechnen. In jeder Iteration wird die Menge der blockierten Reaktionen erweitert. Je größer die Menge der geblockten Reaktionen wird, desto kleiner wird die Menge möglicher EM. Um zu vermeiden, dass keine EM mehr berechnet werden können und außerdem eine ausreichende Diversität der berechneten EM zu gewährleisten, werden Reaktionen mit einer gewissen Wahrscheinlichkeit wieder aus der Menge der blockierten Reaktionen entfernt. In jeder Iteration h wird mit der Wahrscheinlichkeit κ die Reaktion, die zuletzt zu Q_h hinzugefügt wurde, wieder entfernt. Zusätzlich wird mit einer Wahrscheinlichkeit $\lambda \ll \kappa$ die gesamte Menge der blockierten Reaktionen Q_h verworfen und durch $Q_0 = \emptyset$ ersetzt.

Um die Heuristik zu randomisieren und möglichst voneinander unabhängige EM zu bestimmen, wird ein Pool von Knockout-Sets $Q^{(1)}, Q^{(2)}, \dots$ aufrecht erhalten. In jeder Iteration wird ein Knockout-Set zufällig ausgewählt und durch die zuvor beschriebenen Schritte erweitert. Diese Formulierung erlaubt zudem eine einfache Parallelisierung (bei verfügbarer Mehrkern-Computer-Architektur) der Prozedur, bei der Mengen blockierter Reaktionen entwickelt und anschließend die resultierenden linearen Optimierungsprobleme gelöst werden. Hat der verwendete Solver eine Warmstart-Option, so kann die Optimierung zusätzlich beschleunigt werden. Dabei wird die Lösung des vorangegangenen Problems als Basis für eine neue zulässige Lösung genutzt.

Da in jeder Iteration verschiedene Mengen blockierter Reaktionen $Q^{(1)}, Q^{(2)}, \dots$ als Basis zur Berechnung der weiteren EM genutzt werden, können in der Regel neue EM gefunden werden. Die neu hinzugefügten blockierten Reaktionen können zufällig aus der Menge aller Reaktionen des gesamten Netzwerkes gezogen werden. Falls EM mehrfach bestimmt wurden, können diese in einem Nachverarbeitungsschritt wieder entfernt werden.

Für die unten präsentierten Berechnungen werden folgende Werte für die vier Parameter der Methode gewählt:

1. Biomassesyntese als Reaktion t , so dass nur solche EM gefunden werden, die Wachstum ermöglichen.
2. Die Entfernung der zuletzt zur Menge der blockierten Reaktionen hinzugefügten Reaktion erfolgt mit einer Wahrscheinlichkeit von $\kappa = 0.5$.
3. Die Wahrscheinlichkeit, eine gesamte Knockout-Menge zu verwerfen, beträgt $\lambda = 0.002$.
4. Es werden 100 parallele Mengen an Knockout-Mengen unterhalten.

Effizienz des Samplings in genomskaligen Netzwerken

Um die Effizienz des EFMSamplers zu bestimmen, wurden EM für das genomskalige Netzwerk iAF1260 (Feist et al., 2007) berechnet. Dabei wurden für drei verschiedene Bedingungen für jeweils 100 Sekunden EM berechnet; bei der Produktion von i) Succinat, ii) Lycopin oder iii) Biomasse. Mit einem typischen Desktop-PC und drei parallelen Prozessen konnten durchschnittlich 5020 succinatproduzierende EM, 6780 lycopinproduzierende EM und 13475 biomasseproduzierende EM generiert werden. Dies entspricht einer durchschnittlichen Generierungsgeschwindigkeit von 50 bis 135 EM pro Sekunde. Zum Vergleich: Die Berechnung der 100 kürzesten EM mit dem Ansatz von (de Figueiredo et al., 2009) für die drei genannten Szenarien benötigte zwischen ca. vier Minuten (für die Produktion von Succinat und Lycopin) und 18 Minuten (für Biomasse). Damit war diese um mehrere Größenordnungen langsamer als der EFMSampler. Selbst die Berechnung von einer Million biomasseproduzierenden EM mit dem EFMSampler benötigte in etwa nur 85 Minuten.

5.3.3 Analyse der Qualität von Elementarmodenteilmengen

Um die Qualität von mit dem EFMSampler bestimmten EM-Stichproben zu analysieren, wird das oben vorgestellte Zentralstoffwechselnetzwerk verwendet (siehe Unterkapitel 4.1.2). Werden keine weiteren Nebenbedingungen gefordert, umfasst die vollständige Menge für dieses Netzwerk 353196 Elementarmoden. Diese Menge wird für die folgenden Analysen als Grundgesamtheit aller EM verwendet. Es werden 200 verschiedene EM-Teilmengen mit dem EFMSampler berechnet, wobei jeweils 100 Teilmengen 5% ($\mathbb{S}_{5\%}$) bzw. 20% ($\mathbb{S}_{20\%}$) der EM der Grundgesamtheit umfassen.

Die Teilmengen sollen eine möglichst gleichmäßige Abdeckung der Grundgesamtheit der EM und damit des gesamten Lösungsraums gewährleisten. Anhand der folgenden vier Eigenschaften wird diese Anforderung überprüft:

1. *Relative Reaktionshäufigkeiten*
sollten denen der Grundgesamtheit ähnlich sein.
2. *Überdeckung der konvexen Hülle*
sollte die konvexe Hülle der Grundgesamtheit weitgehend überdecken.
3. *Verteilung der EM-Längen*
sollte der Grundgesamtheit ähnlich sein.
4. *Identifizierung von Interventionskandidaten*
sollten denen der Grundgesamtheit entsprechen.

Verteilung der relativen Reaktionshäufigkeiten

Die relativen Reaktionshäufigkeiten der Teilmengen konnten die exakten Werte der Grundgesamtheit sehr gut wiedergeben (Abbildung 5.21). So betrug die durchschnittliche Abweichung (aller Reaktionen aller Teilmengen) für die Teilmengen $\mathbb{S}_{5\%}$ 0.06 und für die Teilmengen $\mathbb{S}_{20\%}$ 0.04 (Tabelle 5.4). Für jede einzelne Stichprobe lag der Pearson-Korrelationskoeffizient für die $\mathbb{S}_{5\%}$ Teilmengen zwischen 0.97 und 0.98 sowie für die $\mathbb{S}_{20\%}$ Teilmengen zwischen 0.98 und 0.99 (Tabelle 5.4).

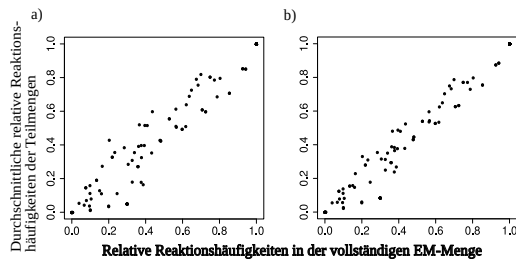


Abbildung 5.21: Durchschnittliche relative Reaktionshäufigkeit der 100 Teilmengen mit a) 5% und b) 20% der EM im Vergleich zur Grundgesamtheit aller EM.

Überdeckung der konvexen Hülle

Um die Überdeckung der konvexen Hülle der Grundgesamtheit zu untersuchen, wurde exemplarisch die Analyse auf die Projektion der EM in eine zweidimensionale Phänotypenebene beschränkt. Hier wird diese durch die Biomassereaktion und die Sauerstoffaufnahme aufgespannt. Da die Teilmengen ausschließlich EM mit obligatorischer Biomassessynthese enthalten, wurden für diese Analyse nur die EM der Grundgesamtheit berücksichtigt, bei denen die Biomassessynthese aktiv ist.

Alle Teilmengen konnten den Großteil des gesamten Lösungsraumes abbilden (Abbildung 5.22). Die durchschnittliche Überdeckung des Lösungsraumes betrug 92.1% mit den $S_{5\%}$ Teilmengen und sogar 99.9% mit den $S_{20\%}$ Teilmengen. Selbst die schlechtesten Überdeckungen konnten 82.3% ($S_{5\%}$) bzw. 99.0% ($S_{20\%}$) des Lösungsraumes abbilden (Tabelle 5.4).

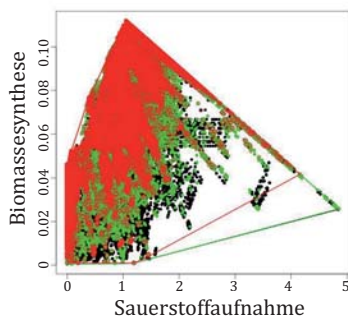


Abbildung 5.22: Überdeckung der konvexen Hüllen der vollständigen Menge aller EM (schwarze Linie und Punkte) und Teilmengen mit Kardinalität von 5% (grün) bzw. 20% (rot) der Grundgesamtheit.

Verteilung der Elementarmodenlängen

Die Längenverteilung der biomasseproduzierenden EM entspricht qualitativ der Verteilung der Grundgesamtheit, wobei eine kleine Abweichung zugunsten kürzerer EM beobachtet werden konnte (Abbildung 5.23). Verursacht wird dies dadurch, dass der EFMSampler neue EM durch das Blockieren zusätzlicher Reaktionen generiert. Dies hat zur Folge, dass längere EM seltener generiert werden, da sie mit einer höheren Wahrscheinlichkeit in späten Iterationen blockiert werden.

Da für biotechnologische Fragestellungen kürzere EM als relevanter angesehen werden können, ist diese Eigenschaft des EFMSamplers keine schwerwiegende Einschränkung. Auch aus biologischer Sicht werden kürzere EM als die Wichtigeren betrachtet, da entlang ihrer Pfade die Investitionen in die Enzymmenge minimiert werden können (de Figueiredo et al., 2009; Fell und Thomas, 1995; Niederberger et al., 1992; Schuetz et al., 2012).

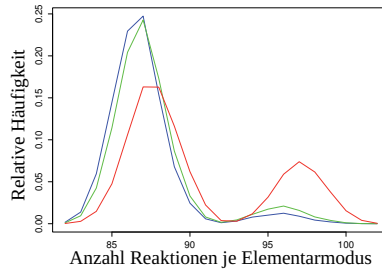


Abbildung 5.23: Längenverteilung der biomasseproduzierenden EM. Rote Linie: Grundgesamtheit der EM, grüne Linie: durchschnittliche Verteilung der EM aus $\mathbb{S}_{20\%}$, blaue Linie: durchschnittliche Verteilung der EM aus $\mathbb{S}_{5\%}$.

Übereinstimmung der CASOP-Maße

Abschließend werden die EM-Teilmengen dahingehend untersucht, ob bei der Anwendung von EM-basierten Methoden vergleichbare Resultate erzielt werden können. Als Anwendungsbeispiel wird hier CASOP verwendet und die Interventionspotentiale Z_2 werden als Vergleichsgrößen herangezogen. Als Produkte werden Succinat und Lycopin gewählt, die jeweils als Vertreter des Zentral- und Sekundärstoffwechsels dienen. Für Succinat wird zusätzlich zwischen aeroben und anaeroben Bedingungen unterschieden, wobei die anaeroben Bedingungen durch alle EM definiert werden, bei denen die Sauerstoffaufnahme-reaktion nicht enthalten ist. Es werden somit drei verschiedene Fallstudien untersucht, für die jeweils die Interventionspotentiale Z_2 für $k = 2$ berechnet werden.

Tabelle 5.4: Qualitätskriterien von jeweils 100 EM-Teilmengen mit 5% ($\mathbb{S}_{5\%}$) bzw. 20% ($\mathbb{S}_{20\%}$) der EM der Grundgesamtheit.

Qualitätskriterium	Teilmengen	
	$\mathbb{S}_{5\%}$	$\mathbb{S}_{20\%}$
	<u>Relative Reaktionshäufigkeiten</u>	
Durchschnittliche Abweichung	0.06	0.04
Spannweite der Abweichung	0.0 - 0.26	0.0 - 0.22
Spannweite der Pearson-Korrelation	0.97 - 0.98	0.98 - 0.99
	<u>Überdeckung der konvexen Hülle</u>	
Durchschnittliche Abdeckung	92.1%	99.9%
Spannweite der Abdeckung	82.3% - 99.9%	99.0% - 100%
	<u>CASOP Interventionspotential Z_2</u>	
Durchschnittliche Abweichung	0.03	0.02
Spannweite der Abweichung	0.0 - 0.26	0.0 - 0.19

Die durchschnittlichen Werte beziehen sich jeweils auf alle Teilmengen. Die Spannweiten berücksichtigen die Werte jeder einzelnen Teilmenge aus $\mathbb{S}_{5\%}$ oder $\mathbb{S}_{20\%}$.

Die Interventionspotentiale der Grundgesamtheit werden sehr gut durch die Teilmengen wiedergegeben. Die durchschnittlichen Abweichungen betragen lediglich 0.03 und 0.02 in den zugehörigen $\mathbb{S}_{5\%}$ bzw. $\mathbb{S}_{20\%}$ Teilmengen (Tabelle 5.4). Zudem können die besten Knockout- und Überexpressionskandidaten auch mit den Teilmengen beider Größen identifiziert werden. Da sowohl Succinat als auch Lycopin in den CASOP-Fallstudien des Kapitels 6.2.1 detailliert betrachtet werden, wird an dieser Stelle auf die Darstellung und Diskussion der identifizierten Kandidaten verzichtet. Es kann jedoch festgehalten werden, dass sowohl die $\mathbb{S}_{20\%}$ - als auch die $\mathbb{S}_{5\%}$ -Teilmengen ausreichend sind, um belastbare CASOP-Maße zu berechnen, mit deren Hilfe Interventionsziele identifiziert werden können.

Zusammenfassend kann gefolgert werden, dass der EFMSampler weitgehend unverzerrte Teilmengen generiert. Diese erfassen die wesentlichen Eigenschaften der vollständigen Menge der EM und erlauben die Anwendung einiger elementarmodenbasierter Verfahren in (genomskaligen) Netzwerken, für die die vollständige Menge der EM nicht berechnet werden kann.

6 Anwendungen der neuen Methoden

Im folgenden Kapitel werden die neu eingeführten Methoden anhand von Beispielen zur Identifizierung von Interventionsstrategien zur Produktion bestimmter Substanzen mit *E. coli* demonstriert. In 6.1 werden die ausgewählte Zielprodukte zunächst vorgestellt. Anschließend werden in 6.2 konkrete Interventionen identifiziert sowie deren Ergebnisse diskutiert.

Zwar liegt der Schwerpunkt dabei auf der Erläuterung der Methoden, dennoch werden einige ausgewählte Ergebnisse detailliert im biologischen Kontext erörtert.

6.1 Betrachtete Produkte und bekannte Interventionsstrategien

Als Beispielprodukte werden Bernsteinsäure und Milchsäure als Repräsentanten von Zielmetaboliten des Zentralstoffwechsels verwendet. Da diese unter physiologischen Bedingungen ausschließlich in ihrer dissoziierten Form vorliegen, werden im Folgenden die entsprechenden Anionen Succinat und Lactat betrachtet. Als Vertreter eines Metaboliten des Sekundärstoffwechsels wird die Produktion von Lycopin untersucht. Für jeden dieser Zielmetabolite gibt es bereits viele experimentelle Studien, so dass auf Basis der verfügbaren Literatur die *in silico* identifizierten Interventionsstrategien zumindest teilweise anhand experimentell verifizierter Interventionen in *E. coli* überprüft werden können. Zudem sind diese Produkte nicht nur von akademischem Interesse, sie stellen ökonomisch vielversprechende Substanzen dar, die zur biotechnologischen Herstellung aus erneuerbaren Rohstoffen geeignet sind (Bozell und Petersen, 2010).

6.1.1 Succinat

Die Literatur zur Succinatproduktion mit *E. coli* ist dermaßen umfangreich, dass eine vollständige Übersicht allein von Übersichtsartikeln dieses Gebietes den Rahmen dieser Arbeit übersteigen würde. Einige der aktuelleren Arbeiten sind die Reviews von (Song und Lee, 2006), (Beauprez et al., 2010), (Cheng et al., 2012) und (Thakker et al., 2012). Diese Arbeiten geben einen generellen Überblick über den aktuellen Stand der biotechnologischen Succinatproduktion. Verschiedene Organismen werden auf ihre Eignung als Succinatproduzenten verglichen und zudem die genetischen Interventionsstrategien sowie die Fermentationsbedingungen der jeweiligen Organismen diskutiert.

Biotechnologische Bedeutung

Succinat hat in den letzten Jahren viel Aufmerksamkeit erhalten, da es als Präkursor für verschiedene Grundchemikalien dient, die zum Beispiel in der Nahrungsmittel- oder Pharmaindustrie verwendet werden (Sauer et al., 2008). Als vielseitig einsetzbare Plattformchemikalie hat es das Potential, andere Grundchemikalien, wie z.B. 1,4-Butanediol zu ersetzen (Bechthold et al., 2008). Die globale Produktionsmenge lag 2006 bei etwa 16.000t (Hermann et al., 2007) und liegt aktuell zwischen 30.000t und 50.000t pro Jahr. Eine Studie (Quelle: Unternehmensberatung CNCC) prognostiziert eine jährliche Steigerung der Produktion von fast 20 % bis ins Jahr 2016. Succinat wird bisher fast ausschließlich petrochemisch hergestellt, da noch kein biotechnologischer Prozess etabliert werden konnte, der ökonomisch mit der petrochemischen Succinatsynthese konkurrieren kann. Dieser Schritt steht jedoch kurz bevor, wie eine Aufzählung der aktuell im Aufbau befindlichen industriellen Anlagen zur biotechnologischen Succinatproduktion belegt (Beauprez et al., 2010). Dabei muss neben der Identifizierung geeigneter Stoffwechselinterventionen auch der Einfluss verschiedener Substrate und der Einfluss der Aufreinigungskosten auf den Gesamtprozess berücksichtigt werden. Um den letzten Schritt hin zur erfolgreichen Kommerzialisierung zu nehmen, müssen zudem klassische biotechnologische Fragestellungen der Prozessentwicklung gelöst werden (Thakker et al., 2012).

Um ökonomisch kompetitiv gegenüber der erdölbasierten Herstellung von Succinat zu sein, muss bei der biotechnologischen Herstellung eine Ausbeute von etwa 0.9 (g Succinat / g Glukose), eine Rate zwischen 1.8 und 2.5 g/l/h und ein Titer von etwa 80 g/l erreicht werden (Beauprez et al., 2010). Bisher ist noch kein Stamm entwickelt worden, der alle drei Anforderungen erfüllt.

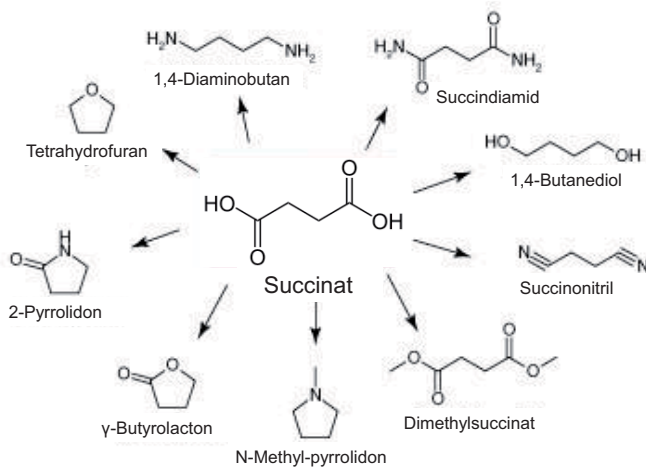


Abbildung 6.1: Succinat als Plattformchemikalie, aus (Sauer et al., 2008).

***In silico* prädizierte und experimentell validierte Interventionen**

Wegen der zuvor erläuterten großen biotechnologischen Bedeutung ist Succinat bereits häufig als Zielmetabolit gewählt worden, um *in silico*-Analysen zur Identifikation von Interventionsstellen zu demonstrieren (Tabelle 6.1). Gemeinsam ist den vorgeschlagenen Interventionsstrategien, dass sie die Ausscheidung von anderen Nebenprodukten unterdrücken. Auch das Zucker-PTS-System wird von verschiedenen Methoden (OptKnock, GDLS, SimOptStrain) als Knockout-Kandidat identifiziert. Unter aeroben Bedingungen wird zudem die Succinat-Dehydrogenase (*sdh*) als Knockout-Kandidat identifiziert.

Tabelle 6.1: *In silico*-Targets für die Produktion von Succinat mit *E. coli*. Δ : Knockout, $\uparrow$: Überexpression, $\downarrow$: Inhibierung

Ansatz	Strategie	Modell	Referenz
OptKnock ¹	$\Delta\{\text{pfl, ldh, adhE}\}$	iJE660	(Burgard et al., 2003)
	$\Delta\{\text{pyk, ack/pta, pts}\}$	iJE660	(Burgard et al., 2003)
	$\Delta\{\text{adhE, ack, pta}\}$		
	$\Delta\{\text{atpA, fum, pta}\}$	iJR904	(Tepper und Shlomi, 2010)
RobustKnock ¹	$\Delta\{\text{ack, fum, ndh}\}$	iJR904	(Tepper und Shlomi, 2010)
OptReg ¹	$\Delta\{\text{pfl}\}, \downarrow \text{adhE}, \uparrow \text{ppc}$		
OptForce ¹	$\downarrow \text{mdh}, \uparrow \text{ppc}, \uparrow \text{cs}$	iAF1260	(Ranganathan et al., 2010)
GDLS ¹	$\Delta\{\text{pts, adhE, rpe}\}$	iAF1260	(Lun et al., 2009)
SimOptStrain ²	$\Delta\{\text{sdhC, pta, eutD}\}$	iAF1260	(Kim et al., 2011)
	$\Delta\{\text{sdhC, gnd, glyA, pntA}\}$	iAF1260	(Kim et al., 2011)
	$\Delta\{\text{cyoA, cydA, lpd, ...}$	iAF1260	(Kim et al., 2011)
	$\text{...ptsH, atpA}\}$		

¹ anaerobe Bedingungen, ² aerobe Bedingungen

Ohne gezielte Interventionen scheidet der *E. coli*-Wildtyp unter aeroben Bedingungen kein und unter anaeroben Bedingungen nur geringe Mengen Succinat aus (Wendisch et al., 2006). Die Interventionen zur Steigerung der Produktion von Succinat in *E. coli* lassen sich in zwei Klassen unterteilen. Zur ersten Gruppe gehören Interventionen, die Schlüsselenzyme der verschiedenen Routen zur Succinat-Synthese verstärken. So konnte die Succinat-Produktion signifikant erhöht werden, indem z.B. die endogene Phosphoenolpyruvat-Carboxylase (*ppc*) (Millard et al., 1996) oder die Pyruvat-Carboxylase (*pyc*) (Vemuri et al., 2002b,a) überexprimiert wurde. Auch die Überexpression der Fumarat-Reduktase (*frdABCD*) führte zu einer deutlichen Steigerung der Succinat-Synthese (Goldberg et al., 1983; Wang et al., 1998). Des Weiteren hat die Expression einiger exogener Enzyme einen positiven Einfluss auf die Succinat-Produktion, darunter z.B. die Phosphoenolpyruvat-Carboxykinase von *Actinobacillus succinogenes* (Kim et al., 2004) sowie die Pyruvat-Carboxylase von *Lactococcus lactis* (Lin et al., 2005b).

Zur zweiten Klasse gehören Interventionen, die die Pathways der Nebenprodukte eliminieren bzw. inhibieren. So wurden Mutanten konstruiert, in denen alle alternativen Fermentationswege (Lactat: *ldhA* und *mgsA*, Ethanol: *adhE*, Acetat: *ackA-pta* und *poxB*, Formiat: *pflB* und *focA*) eliminiert wurden (Jantama et al., 2008a,b). Weitere erfolgreiche Strategien beinhalteten Knockouts der Succinat-Dehydrogenase *sdh*, der Glukose-Phosphotransferase *ptsG*, der Ketoglutarat-Dehydrogenase *sdhAB* sowie des Repressors der Isocitrat-Lyase *iclR*. Ein umfassender Überblick spezieller *E. coli* Stämme zur Succinatproduktion unter verschiedenen Kultivierungsbedingungen ist in (Beauprez et al., 2010) und (Cheng et al., 2012) gegeben. Eine Auswahl experimentell validierter Interventionen ist in Abbildung 6.2 dargestellt.

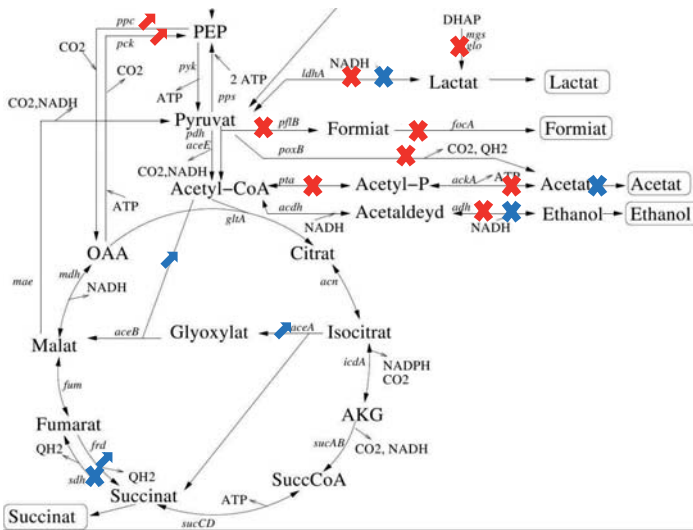
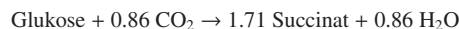


Abbildung 6.2: Experimentell validierte Interventionen zur Succinat-Produktion. Kreuze bzw. Pfeile: Knockouts bzw. Überexpressionen, rote bzw. blaue Symbole: anaerobe bzw. aerobe Bedingungen.

Optimale Stöchiometrie der Succinatproduktion

Anhand einer Analyse der optimalen Gesamtreaktion der Succinatproduktion können erste Aussagen bezüglich der Co-Faktorbedeutung getroffen werden. Diese Analysen basieren ausschließlich auf der stöchiometrischen Bilanz sowie der Elektronenbilanz. Die berechneten Produktausbeuten sind daher obere Schranken für alle spezifischen Stoffwechselnetzwerke. Die optimale Umsetzung von Glukose zu Succinat ist gegeben durch:



Ohne CO_2 -Aufnahme ist die Succinat-Ausbeute auf 1.5 mol pro mol Glukose beschränkt. Der Vorteil einer CO_2 -Supplementierung für optimale Succinat-Produktion ist somit leicht ersichtlich. Die maximale Ausbeute wird zudem durch die verfügbare Reduktionskraft

beschränkt. Pro mol Glukose stehen 24 freie Elektronen zur Verfügung (der Reduktionsgrad von Glukose, berechnet als Anzahl freier Elektronen pro Kohlenstoffatom, ist 4). Der Reduktionsgrad von Succinat ist 3,5, so dass maximal $(4 \cdot 6) / (3,5 \cdot 4) = 1,71$ mol Succinat gebildet werden können. Kürzlich vorgestellte Studien (Lin et al., 2005a; Sánchez et al., 2005; Lu et al., 2009) bestätigten die Prädiktionen der positiven Effekte von CO₂ und NADH. So konnte unter anderem gezeigt werden, dass für einen *E. coli* Stamm (mit Deletion von *pfl*, *ptsG* und *ldh*) erhöhte CO₂-Konzentrationen unter anaeroben Bedingungen zu einem dramatischen Anstieg der Succinat-Produktivität von 1,9 mg/(gDW·h) zu 225 mg/(gDW·h) führen. Zudem ist diese Beobachtung an eine höhere Aktivität von NAD(P)H-produzierenden Pfaden gekoppelt. In weiteren Studien konnte gezeigt werden, dass mit einer 100%-igen CO₂-Atmosphäre eine gesteigerte Succinat-Produktion einhergeht (Sánchez et al., 2006; Jantama et al., 2008a).

6.1.2 Lactat

Biotechnologische Bedeutung

Die weltweite Produktion von Lactat nahm in den vergangenen Jahren stetig zu, wobei die aktuelle jährliche Steigerungsrate etwa 10% beträgt. So verdreifachte sich die weltweite Produktion in den Jahren von 1990 bis 2005 von 40.000t pro Jahr auf 120.000t pro Jahr (Datta und Henry, 2006), wovon bereits 90% biotechnologisch durch Fermentation mit Milchsäurebakterien gewonnen wurde (Hofvendahl und Hahn-Hägerdal, 2000). In den darauf folgenden fünf Jahren stieg die Produktion erneut um mehr als das Doppelte auf 250.000t pro Jahr (Soetaert und Vandamme, 2010) und liegt im Moment (2014) bei etwa 300.000t bis 400.000t pro Jahr. Bis zum Jahr 2020 ist ein Anstieg auf 950.000t pro Jahr prognostiziert (Studie nova-Institut).

Wurde Milchsäure früher vor allem in der Lebensmittelindustrie verwendet, nimmt seine Bedeutung seit einigen Jahren als Massenchemikalie stetig zu (Datta et al., 1995; Hofvendahl und Hahn-Hägerdal, 2000). Die technologischen Möglichkeiten und Beschränkungen möglicher Anwendungsfelder von Lactat werden in (Sauer et al., 2008) und (Datta und Henry, 2006) ausführlich diskutiert. Zudem wird Lactat als Monomer für die Synthese von Poly-Lactic-Acid (PLA) - dem vielversprechendsten biologisch abbaubaren Plastik - benötigt. PLA wird auf Grund seiner mechanischen Eigenschaften (z.B. Elastizität, Festigkeit) das Potenzial zugesprochen, erdölbasierte Kunststoffe wie z.B. PET (Polyethylen Terephthalat) zu ersetzen. Die Steigerung der weltweiten Kapazität der PLA-Produktion bestätigt diese Entwicklung (Abbildung 6.3).

In silico prädizierte und experimentell validierte Interventionen

Naheliegende Interventionsziele sind die Knockouts der alternativen Fermentationswege zur Produktion von Acetat oder Ethanol. Weitere *in silico* Kandidaten sind die Deletion der Phosphofructokinase (*pfk*) oder der Glucokinase (*glk*) (Tab. 6.2).

Umfassende Übersichtsartikel Lactat produzierender *E. coli*-Stämme sind in (Wendisch et al., 2006) und (Yu et al., 2011) gegeben. Die durchgeführten Eingriffe lassen sich grundsätzlich in zwei Kategorien unterscheiden. Eine Klasse beschäftigt sich mit Eingriffen zur Herstellung von optisch reinem D- oder L-Laktat. Die zweite Kategorie von Eingriffen befasst sich mit der Steigerung der Lactat-Produktion, deren wichtigste Resultate hier kurz diskutiert werden.

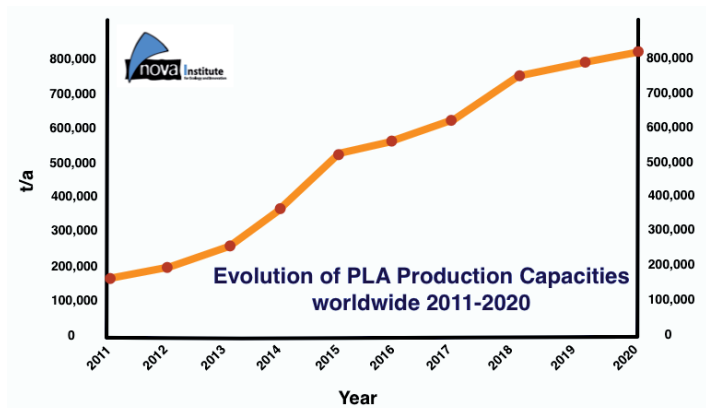


Abbildung 6.3: Prädierte Entwicklung der weltweiten Kapazität der PLA-Produktion.

Die Inaktivierung der Fumarat-Reduktase *frdABCD*, der Alkohol-Dehydrogenase *adhE*, der Pyruvat-Formiat-Lyase *pflB* und der Acetat-Kinase *ackA* bewirken eine Steigerung der Lactat-Ausbeute (Zhou et al., 2003, 2005). Kombinationen von Deletionen der Gene *adhE*, *pta/ackA*, *glk* und *pfkA* wurden von (Zhou et al., 2005) und (Hua et al., 2006) experimentell validiert. Die Repression von Ethanol und Acetat produzierenden Pathways sowie der Knockout von *glk* führt zu gesteigerter Lactat-Produktion. Die Deletion von *pfkA* hingegen verursacht eine intrazelluläre Akkumulation von Fructose-6-Phosphat, welches einen inhibierten Glukosetransport zur Folge hat (Fong et al., 2005). In einem aktuellem Patent wurde ein *E. coli* Stamm konstruiert, bei dem zunächst Gene von Nebenpathways deletiert wurden und anschließend mittels adaptiver Evolution eine Selektion der Mutanten mit verbesserter Laktat-Produktion stattfand. Es konnte ein Stamm isoliert werden, mit dem eine Ausbeute von etwa 90% und eine absolute Produktivität von etwa 2 g/l/h erreicht werden kann (Zhou et al., 2013). Abbildung 6.4 gibt einen Überblick experimentell erfolgreich validierter Interventionen.

Tabelle 6.2: *In silico* Targets für die Produktion von Lactat mit *E. coli*. Δ : Knockout

Ansatz	Strategie	Modell	Referenz
OptKnock ¹	Δ atpA, pdh, pfl	iJR904	(Tepper und Shlomi, 2010)
	Δ ackA/pta, adhE	iJE660	(Burgard et al., 2003)
	Δ ackA/pta, pfk/fba	iJE660	(Burgard et al., 2003)
	Δ ackA/pta, pfk/fba, adhE, glk	iJE660	(Burgard et al., 2003)
RobustKnock ¹	Δ adhE, ack, eno	iJR904	(Tepper und Shlomi, 2010)

¹ anaerobe Bedingungen

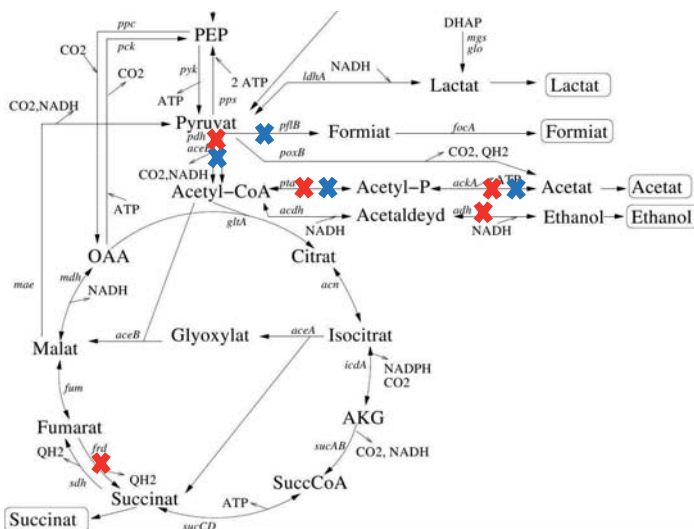
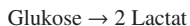


Abbildung 6.4: Experimentell validierte Interventionen zur Lactat-Produktion. Kreuze bzw. Pfeile: Knockouts bzw. Überexpressionen, rote bzw. blaue Symbole: anaerobe bzw. aerobe Bedingungen.

Optimale Stöchiometrie der Lactatproduktion

Die Stöchiometrie der optimalen Umsetzung von Glukose zu Lactat ist gegeben durch:



Da bei der Lactat-Synthese kein CO_2 fixiert werden kann, ist eine externe Supplementierung mit CO_2 aus stöchiometrischer Sicht nicht notwendig. Zudem ist der Reduktionsgrad sowohl von Glukose als auch von Lactat vier, so dass bei ausbeuteoptimalen Flussverteilungen kein Überschuss oder Mangel an Reduktionskraft vorliegt.

6.1.3 Lycopin

Lycopin ist ein wichtiger Vertreter der Carotinoide. Letztere sind natürlich vorkommende Pigmente (Lycopin ist der rote Farbstoff der Tomate), die zur Unterklasse der Tetraterpene gehören. Für die Biosynthese aller Terpene sind Dimethylallylpyrophosphat (DMAPP) und das Doppelbindungsisomer Isopentenylpyrophosphat (IPP) gemeinsame Präkursoren. Beide C5-Metabolite basieren auf Isopren und werden in weiteren verzweigten Reaktionsschritten zu den verschiedenen Klassen von Terpenen verstoffwechselt. *E. coli* besitzt den endogenen Deoxy Xylulose Phosphat (DXP)-Pathway, um diese Präkursoren (IPP und DMAPP) zu synthetisieren (Abb. 6.5). Ein zweiter Syntheseweg, der Mevalonat-Weg, kommt ursprünglich in Pflanzen und eukaryotischen Mikroorganismen wie z.B. *S. cerevisiae* vor. Eine ausführliche Diskussion der optimalen Flussverteilung und geeigneter Überexpressions- und Knockout-Kandidaten zur Überproduktion von IPP mit *E. coli* und *S. cerevisiae* findet sich z.B. in (Gruchattka et al., 2013).

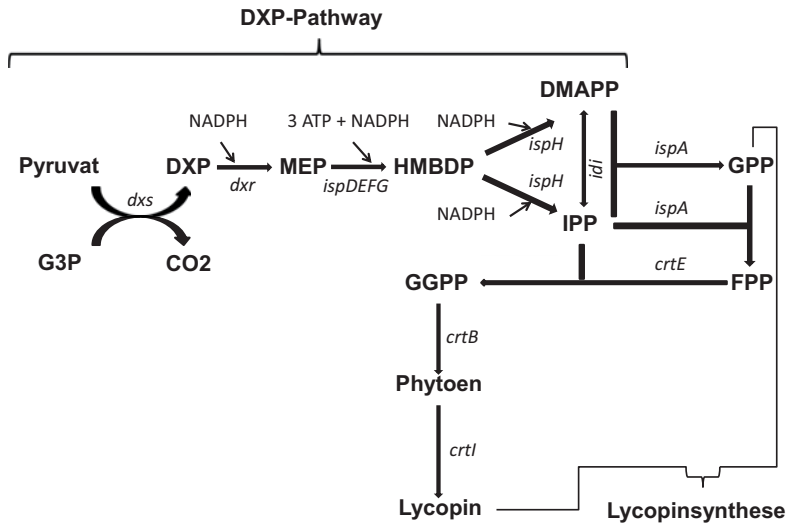


Abbildung 6.5: Reaktionen des DXP-Pathways und der Synthese von Lycopin. Für die verwendeten

Biotechnologische Bedeutung

Terpene sind eine Klasse von Produkten mit wichtigen medizinischen und industriellen Anwendungen wie z.B. Artemisinin (Wirkstoff in Malaria-Medikamenten) oder Paclitaxel (Wirkstoff in Krebs-Medikamenten). Weitere Anwendungen haben sie z.B. in Kosmetika und als Nahrungsergänzungsmittel. Überdies sind Carotinoide hervorragend als Modell- bzw. Testprodukte geeignet, da ihre Konzentration (und damit Produktion) durch kolorimetrische Methoden sehr einfach bestimmt werden kann.

In silico prädiizierte und experimentell validierte Interventionen

Die bisher in der Literatur beschriebenen Interventionsstrategien konzentrieren sich hauptsächlich auf die Überexpression der limitierenden Enzyme wie z.B. *dxs*, *idi*, *dxr*, *ispD* und *ispF*. Die Justierung der optimalen Expressionslevel der Enzyme des DXP-Pathways und weiterer heterologer Überexpressionen stellte sich dabei als besonders erfolgreich heraus (Wang et al., 2009). Weitere Studien zeigen, dass sich durch die Expression des heterologen Mevalonat-Weges in *E. coli* eine Steigerung der Terpenoid-Synthese erreichen lässt (Yoon et al., 2006). Eine Vielzahl diverser Interventionsstrategien wurde bereits *in silico* identifiziert und experimentell überprüft; eine Auswahl ist in Tabelle 6.3 gegeben.

Weitere komplexe Interventionsstrategien mit Eingriffen in den Sekundärstoffwechsel wurden systematisch experimentell überprüft (Alper und Stephanopoulos, 2008; Alper et al., 2005c; Jin und Stephanopoulos, 2007). Nur wenige Studien betrachteten bisher

Eingriffe in den Zentralstoffwechsel zur Steigerung der Lycopinproduktion. In (Vadali et al., 2005) bewirkten Knockouts von *ack/pta* und der NADH-Dehydrogenase (*nuo*) eine Steigerung der Lycopinproduktion gegenüber dem Wildtyp. Die Ausscheidung weiterer alternativer Nebenprodukte sowie Deletionen von *pykF*, *zwf* und *maeB* haben sich zur Produktion zweier Carotinoide als vorteilhaft erwiesen (Unrean et al., 2010).

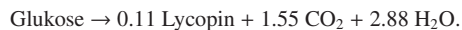
Tabelle 6.3: *In silico* Targets für die Produktion von Lycopin mit *E. coli*. Diese Targets wurden zudem experimentell überprüft (siehe Referenzen) Δ : Knockout

Ansatz	Strategie	Modell	Referenz
FSEOF ¹	$\uparrow$ { <i>pfkA</i> , <i>pgi</i> , <i>fba</i> , <i>tpi</i> <i>mdh</i> , DXP pathway}	EcoMBEL979	(Choi et al., 2010)
MOMA ¹	Δ { <i>gdhA</i> , <i>icdA</i> , <i>cyoA</i> , <i>gpmAB</i> }	EcoMBEL979	(Choi et al., 2010)
FBA mit MOMA ¹	Δ { <i>gdhA</i> , <i>cyoA</i> , <i>gpmAB</i> , <i>ppc</i> , <i>eno</i> , <i>glyA</i> , <i>aceE</i> }	iJR904	(Alper et al., 2005b)
	Δ { <i>gdhA</i> , <i>aceE</i> , <i>fdhF</i> }	iJR904	(Alper et al., 2005b)
MMF ²	Δ { <i>ldhA</i> , <i>frdA</i> , <i>poxB</i> , <i>pta</i> , <i>adhE</i> , <i>pykF</i> , <i>zwf</i> , <i>maeB</i> }	ZSW	(Unrean et al., 2010)

¹ aerobe Bedingungen, ² Zielmetabolite sind zwei verschiedene Carotinoide (nicht Lycopin)

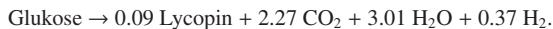
Optimale Stöchiometrie der Lycopinproduktion

Die optimale Stöchiometrie der Umsetzung von Glukose zu Lycopin (allein auf der Kohlenstoff- und Redoxbilanz basierend) ist gegeben durch:

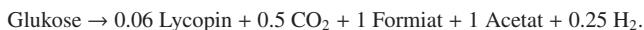


Da unter optimalen Bedingungen CO_2 gebildet wird, ist eine externe Supplementierung mit CO_2 aus stöchiometrischer Sicht nicht notwendig. Die maximale Ausbeute wird auch hier durch die Elektronenbilanz beschränkt. Eine Quelle zusätzlicher Elektronen würde höhere Ausbeuten (bis zu 0.125 mol Lycopin/ mol Glukose) ermöglichen.

Unter den stöchiometrischen Zwängen des Stoffwechsels von *E. coli* ergeben sich reduzierte Lycopinausbeuten. Unter aeroben Bedingungen ist die optimale Lycopinsynthese mit *E. coli* gegeben durch:



Bei anaeroben Bedingungen ergibt sich entsprechend:



Diese reduzierten optimalen Ausbeuten ergeben sich aus dem ATP-Bedarf der Lycopinsynthese. Der ATP und Co-Faktorbedarf der Lycopinsynthese im Stoffwechselnetzwerk von *E. coli* wird daher im nächsten Unterkapitel detailliert diskutiert.

6.2 Berechnung von Interventionsstrategien

Für die vorgestellten Produkte werden im Folgenden Interventionsstrategien für deren Produktion in *E. coli* mit Hilfe der neuen Methoden, CASOP und cMCS, identifiziert. Dafür wird das in Kapitel 4.1.2 vorgestellte Modell des Zentralstoffwechsels verwendet. Es werden sowohl aerobe (aer) als auch anaerobe (an) Produktionsbedingungen betrachtet. Eine Stärke sowohl der CASOP- als auch der cMCS-Methodik ist die Möglichkeit, Interventionsziele zu identifizieren, mit denen verschiedene Strategieziele wie Ausbeute oder Produktivität verfolgt werden können. Um dies zu demonstrieren, wird für jedes Produkt sowohl eine Interventionsstrategie mit Schwerpunkt auf Ausbeuteoptimalität (Y) als auch eine Strategie zur Steigerung der Produktivität (R) identifiziert. Somit werden für jedes Produkt durch die Kombination aus Umweltbedingung und Strategieziel vier verschiedene Szenarien untersucht:

§Bedingung
Produkt,Ziel, Bedingung $\in \{\text{aerob, anaerob}\}$, Ziel $\in \{Y, R\}$, Produkt $\in \{\text{Succ, Lac, Lyco}\}$

Folgende Annahmen werden dabei für die jeweiligen Szenarien getroffen:

1. Anaerobe Nebenbedingungen werden durch die Beschränkung auf die Menge der EM abgebildet, bei denen die Sauerstoffaufnahme Null ist.
2. Da der Wildtyp von *E. coli* kein Lycopin ausscheidet, wird für die Succinat- und Lactatszenarien die Lycopin-Synthese als inaktiv gesetzt.
3. Für alle Szenarien wird eine CO_2 -Verfügbarkeit angenommen.

6.2.1 Anwendung von CASOP

Im Folgenden werden Interventionsziele mit CASOP identifiziert. Für jedes Produkt werden zunächst die Interventionspotentiale Z_2 aller Reaktionen berechnet und geeignete Eingriffe identifiziert. Anschließend wird die Bedeutung der Co-Faktoren für die Produktsynthese analysiert. Die Bewertung der Relevanz der CO_2 -Aufnahme sowie CO_2 - und ATP-Ausscheidung kann durch die im Netzwerk enthaltenen Austauschreaktionen (für CO_2) bzw. die *ATPdrain*-Reaktion erfolgen. Zusätzlich werden für ATP und NADH künstliche Quellen definiert und separat analysiert. Zudem werden für jedes Produkt einzelne Interventionsstrategien iterativ bestimmt, die sich aus drei Knockout- und einem Überexpressionskandidaten zusammensetzen.

Formulierung der Strategieziele als CASOP-Probleme

Wie in Abschnitt 5.1 beschrieben, wird für jedes der drei Produkte ein künstlicher Metabolit ∇ definiert, der sich aus dem jeweiligen Produkt und der Biomasse zusammensetzt. Dabei werden jeweils zwei Zusammensetzungen für ∇ mit verschiedenen relativen Anteilen des Produktes definiert. Um den Phänotypen des Wildtypen abzubilden (der Ausgangssituation), wird der Massenanteil des Produktes auf 0% ($\gamma = 0$) gesetzt. Für den angestrebten Produktionsphänotypen wird der Massenanteil des Produktes an der Biomasse auf 90% gesetzt ($\gamma = 0.9$). Um die Rangfolge der Kandidaten zu bestimmen, wird das Interventionspotential Z_2 berechnet (der Unterschied der Relevanzen zwischen der

Ausgangs- und Produktionssituation). Für Eingriffe, die Produktivität zum Strategieziel haben, wird $k = 0$ gewählt. Um die optimale Ausbeute als Strategieziel stärker zu gewichten, werden zusätzlich $k = 2$ und $k = 10$ verwendet. Bei der Identifizierung multipler Interventionsstrategien wird für produktivitätsorientierte ($\mathbb{S}_{\mathbb{R}}$) bzw. ausbeuteorientierte Szenarien $\mathbb{S}_Y$ entsprechend $k = 0$ und $k = 2$ gewählt. Für die Bewertung der künstlichen ATP- und NADH-Quellen wird $k \in \{0, 1, 2, 5, 10, 20, 100\}$ verwendet.

Ergebnisse für Succinat

Die Interventionspotentiale ausgewählter Reaktionen sind in Tabelle 6.4 gegeben. Für den Großteil der Reaktionen führen größere k zu größeren (absoluten) Interventionspotentialen. Die deutlichsten Knockout-Kandidaten sind *ptsG* (Glc_PTS_up), die Ausscheidungen von Ethanol, Formiat und Acetat sowie die Succinat-Dehydrogenase (*sdh*, bzw. Succ::Fum). Diese Knockouts wurden alle durch experimentelle Studien bereits erfolgreich validiert (6.1.1). Die *sdh* wird hier auch unter anaeroben Bedingungen vorgeschlagen (im Gegensatz zu den in der Literatur vorgeschlagenen Mutanten), da sie im Modell nicht als inaktiv gesetzt wurde (entsprechende regulatorische Nebenbedingungen werden hier nicht berücksichtigt). Die in der Praxis meist verwendeten anaeroben Fermentationsbedingungen werden durch die Identifizierung der Sauerstoffaufnahme (O2_up) als Knockout-Kandidat (unter aeroben Bedingungen) bestätigt. Auch die Überexpressionskandidaten mit den höchsten Interventionspotentialen, *frdA* (Fum::Succ), die CO₂-Aufnahme, *ppc* (PEP::OxA) sowie (unter aeroben Bedingungen) die Reaktionen des Glyoxylat-Wegs (ICit::Glyox, Glyox::Mal), wurden experimentell bereits erfolgreich validiert. Weitere Überexpressionskandidaten, für die bisher keine experimentellen Bestätigungen vorliegen, sind *ppsA* (Pyr::PEP), *mgsA* (DHAP::Lac) und die Reaktion ATPdrain. Diese Kandidaten lassen sich im Zusammenhang mit dem ATP-Stoffwechsel verstehen und werden weiter unten näher erläutert. Die Identifizierung von *fum* (Fum::Mal) als Knockout-Kandidat ist hier den vernachlässigten regulatorischen Nebenbedingungen (wie z.B. Inaktivierung einiger Reaktionen des TCA unter anaeroben Bedingungen) geschuldet, so dass dessen Implementierung in der Praxis keine Rolle spielt. Eine ausführlichere Analyse der Auswirkungen verschiedener, bekannter regulatorischer Nebenbedingungen auf die identifizierten Interventionen wird in (Hädicke und Klamt, 2010) gegeben.

Einfluss von CO₂, NADH und ATP auf die Succinatsynthese

Die Relevanz der CO₂- Aufnahme- und Ausscheidereaktion durch CASOP stimmt mit der Analyse der optimalen stöchiometrischen Flussverteilung überein. So wird die CO₂-Aufnahmereaktion als „Überexpression“ und die CO₂-Ausscheidung als Knockout-Kandidat identifiziert (Tabelle 6.4). Die artifizielle NADH-Aufnahmereaktion wird als Überexpressionskandidat identifiziert (Abbildung 6.6), was ebenfalls in Übereinstimmung mit der vorangegangenen Analyse ist. Die Bewertung deutet darauf hin, dass der Bedarf an zusätzlicher Reduktionskraft besonders für optimale Ausbeuten und ein hohes Succinat zu Biomasseverhältnis gegeben ist. Die Aufnahme von zusätzlichem ATP wird als Knockout klassifiziert. Entsprechend wird die ATPdrain Reaktion als Überexpression vorgeschlagen (Tabelle 6.5). Beide Interventionen werden umso klarer identifiziert, desto stärker der Fokus auf der Ausbeuteoptimalität liegt. Eine Erklärung dafür liefert erneut die ausbeuteoptimale Flussverteilung. Diese erlaubt es, ATP im Überschuss zu generieren, so dass

Tabelle 6.4: CASOP Z₂-Interventionspotentiale für die Produktion von Succinat. Grün hinterlegt: Übereinstimmung mit experimentell bestätigten Interventionen. Grau hinterlegt: experimentell bestätigte Intervention, kein CASOP-Kandidat. Rot hinterlegt: interessante, neu identifizierte Intervention.

Pfad	Reaktion	aerob			anaerob		
		0	2	10	0	2	10
Aufnahme	O ₂ _up	-0,1893	-0,2081	-0,3601	-	-	-
	CO ₂ _up	0,1051	0,184	0,5182	0,1904	0,4331	0,8878
	Glc_PTS_up	-0,0599	-0,0591	-0,1151	-0,1527	-0,2678	-0,3488
	Glc_ATP_up	0,0522	0,0312	0,0865	0,1457	0,2392	0,5131
EMP	G6P::F6P	0,0872	0,1398	0,1807	0,0724	0,0716	0,0091
	F6P::F16P/F16P::T3P	0,1056	0,1672	0,1547	0,052	0,0735	0,0383
	DHAP::G3P	0,0201	0,0348	0,1096	0,1129	0,121	0,0506
	G3P::DPG/DPG::3PG	0,0145	0,0187	0,0207	0	0	0
	3PG::2PG/2PG::PEP	0,0164	0,0269	0,0336	0	0	0
	PEP::PYR	0,0332	0,0274	-0,0123	0,0714	-0,0279	-0,278
	Pyr::PEP	0,0826	0,103	0,1449	0,0682	0,1269	0,284
PPP	G6P::PGLac/PGLac::PGLuc	-0,029	-0,0473	-0,0504	-0,023	-0,0274	0,0757
	PGLuc::R15P	-0,0271	-0,0615	-0,0822	0,0099	0,0599	0,2142
	R15P::X5P	-0,0094	-0,0265	-0,0615	-0,0241	-0,0397	-0,0655
	R15P::R5P	0	0	0	0	0	0
	Transket1/Transaldo	-0,0326	-0,0533	-0,0763	-0,0446	-0,0438	-0,0432
	Transket2	0,0258	0,0377	0,0054	0,0013	-0,0112	-0,078
EDD	2KD6PG::PyrG3P	-0,1065	-0,147	-0,1438	-0,1141	-0,1928	-0,0781
TCA	PYR::AcCoA	-0,0575	-0,0825	-0,1206	0,046	0,2255	0,4286
	AcCoA::Cit	0	0	0	0	0	0
	Cit::ICit/ICit::alKG	0	0	0	0	0	0
	alKG::SuccCoA	-0,0003	-0,0244	-0,0617	0,052	0,0745	0,2175
	SuccCoA::Succ	0,0159	0,0031	-0,0261	0,0103	-0,0021	-0,0134
	Succ::Fum	-0,7604	-0,8462	-0,932	-0,7362	-0,7388	-0,5988
	Fum::Succ	0,6799	0,772	0,9685	0,8286	0,9559	1
	Fum::Mal	-0,2687	-0,1924	-0,0281	-0,2461	-0,0775	0
	Mal::Ox_A_NADH	-0,0926	-0,0791	0,051	-0,0479	0,1206	0,3626
ANA	Mal::PyrNADH	-0,0856	-0,1371	-0,1965	-0,1197	-0,2283	-0,2533
	Ox_A::PEP	-0,0104	-0,0181	0,0227	0,0342	0,0478	0,1029
	PEP::Ox_A	0,3371	0,4099	0,4406	0,3572	0,3656	0,177
GlyOx	ICit::Glyox/Glyox::Mal	0,1911	0,0735	0,0141	0,1867	0,0637	0,2212
Ferm	AcCoA::Adh/Eth_ex	0,0874	0,0901	0,0381	-0,0452	-0,2708	-0,8539
	AcCoA::AcP/AcP::Ac	-0,0156	-0,0071	0,0108	-0,0016	-0,019	-0,2706
	Pyr::Form/Form_ex	0,0934	0,1332	0,1164	-0,1519	-0,3654	-0,5562
	Pyr::Lac	-0,0415	0,0064	0,0145	-0,0406	-0,0024	0,0551
	Pyr::Ac	0,0459	0,0477	0,0888	0,2016	0,2071	0,0802
	DHAP::Lac	0,1191	0,0515	0,0023	0,1741	0,1373	0,1619
	Lac_ex	0,0089	-0,0254	0,0009	-0,0393	-0,0344	-0,0148
	Ac_ex	-0,0793	-0,0676	-0,0277	-0,0421	-0,0786	-0,3473
	Succ_Ex	1	1	1	1	1	1
ETC/AUX	NADHDehydro	0,0242	-0,01	-0,0007	-0,0934	0,0705	0,3212
	Oxidase	-0,1893	-0,2081	-0,3601	0	0	0
	TransHydro	0,1842	0,2594	0,0446	-0,0935	-0,1295	-0,1342
	ATPSynth	-0,056	-0,0519	-0,0134	0,0286	0,1458	0,2493
	ATPdrain	0,1039	0,1185	0,1194	0,1407	0,14	0,1153
	CO ₂ _ex	-0,4004	-0,5433	-0,8632	-0,506	-0,6856	-0,6552

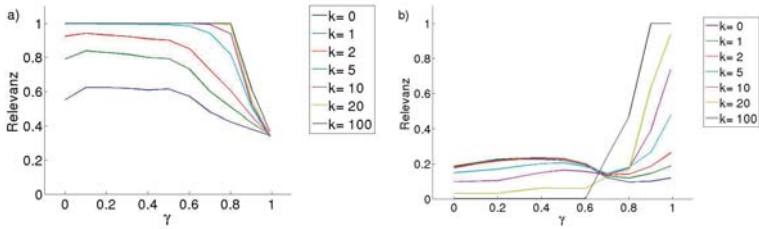


Abbildung 6.6: a) Relevanz einer artifiziiellen ATP- und b) NADH-Aufnahme für Succinatproduktion unter anaeroben Bedingungen für verschiedene k .

bis zu 1.38 mol ATP pro mol Glukose bei der Succinatsynthese gewonnen werden können. Dies bedeutet zum Einen, dass die optimale Succinatausbeute nicht durch den ATP-Bedarf beschränkt ist. Des Weiteren kann somit bei hohem Succinat- zu Biomasseverhältnis ein ATP-Überschuss vorliegen, der über geeignete Senken ausgeglichen werden muss. In diesem Zusammenhang können nun auch die identifizierten Überexpressionskandidaten verstanden werden. Wird Glukose über den Methylglyoxalweg (Reaktion *mgsA*) zu Pyruvat metabolisiert, so wird dabei im Gegensatz zur Glykolyse kein ATP gewonnen und ATP-unabhängige Reaktionswege favorisiert. Ebenso wird durch eine Überexpression von *ppsA* ein ATP verbrauchender Zyklus implementiert, der analog wie ein erhöhter ATPdrain Fluss als ATP-Senke fungieren kann.

Multiple Eingriffe zur Steigerung der Succinatsynthese

Um konkrete multiple Interventionsstrategien zu identifizieren, wird die CASOP-Methodik iterativ angewendet. Die besten drei Knockout-Kandidaten jeder Iteration sind in Tabelle 6.5 dargestellt. Die besten Kandidaten der ersten Iteration sind über alle Szenarien hinweg sehr ähnlich. So wird die *sdh* für alle Szenarien als bester Knockout-Kandidat identifiziert. In den nachfolgenden Iterationen werden für anaerobe Bedingungen die Knockouts von *sucCD* und *fum* bzw. *pfl* und der Acetatausscheidung identifiziert. Unter anaeroben Bedingungen wird für Szenario $S_{\text{Succ,R}}^{\text{an}}$ nach den drei vorgeschlagenen Knockouts eine Kopplung von $\gamma_{\min} = 0.15$ erreicht. Für Szenario $S_{\text{Succ,Y}}^{\text{an}}$ ist bereits nach dem zweiten Knockout eine Kopplung von 0.06 gewährleistet, die sich mit dem dritten Knockout auf 0.45 erhöht. Auch unter aeroben Bedingungen kann eine Kopplung für Szenario $S_{\text{Succ,R}}^{\text{aer}}$ von 0.15 nach den drei Knockouts erreicht werden. Für Szenario $S_{\text{Succ,Y}}^{\text{aer}}$ ist nach zwei Knockouts eine Kopplung von 0.06 erreicht, die durch den dritten Knockout nicht erhöht wird. Als Überexpressionskandidaten werden für aerobe Bedingungen in Übereinstimmung mit experimentellen Ergebnissen die Reaktionen des Glyoxylat-Shunts identifiziert. Wie zuvor werden *mgs*, *ppsA* und ATPdrain für mehrere Szenarien als Überexpressionskandidaten identifiziert. Diese Interventionen wurden bisher weder von einer anderen *in silico*-Methode identifiziert noch gezielt experimentell überprüft. Die experimentelle Überexpression von *ppsA* ist daher Gegenstand von Kapitel 8.

Tabelle 6.5: CASOP Z₂-Rating für die jeweils Top 3 Knockout- (Iteration 1-3) und Überexpressionskandidaten (Iteration 4). Grün hinterlegt: Übereinstimmung mit experimentell bestätigten Interventionen. Fett: in der jeweiligen Iteration ausgewählter *in silico* Eingriff.

Iteration	$\bar{S}_{\text{Succ,R}}^{\text{an}}$	$\bar{S}_{\text{Succ,Y}}^{\text{an}}$	$\bar{S}_{\text{Succ,R}}^{\text{aer}}$	$\bar{S}_{\text{Succ,Y}}^{\text{aer}}$
Erste	<i>sdh</i> (-0.73)	<i>sdh</i> (-0.74)	<i>sdh</i> (-0.76)	<i>sdh</i> (-0.85)
	<i>fum</i> (-0.25)	<i>pfl</i> (-0.37)	<i>fum</i> (-0.27)	<i>fum</i> (-0.19)
	<i>pts</i> (-0.15)	<i>adhE</i> (-0.27)	<i>eda</i> (-0.11)	<i>eda</i> (-0.15)
	<i>sucCD</i> (-0.36)	<i>sucCD</i> (-0.37)	<i>sucCD</i> (-0.37)	<i>sucCD</i> (-0.37)
Zweite	<i>fum</i> (-0.26)	<i>pfl</i> (-0.37)	<i>fum</i> (-0.30)	Acetat ex (-0.31)
	<i>mae</i> (-0.19)	<i>mae</i> (-0.33)	Acetat ex (-0.26)	<i>mae</i> (-0.22)
	<i>fum</i> (-0.25)	<i>pfl</i> (-0.39)	<i>fum</i> (-0.30)	Acetat ex (-0.26)
Dritte	<i>mae</i> (-0.20)	<i>adhE</i> (-0.34)	Acetat ex (-0.19)	<i>mae</i> (-0.22)
	<i>pfl</i> (-0.17)	<i>mae</i> (-0.33)	<i>mae</i> (-0.15)	<i>fum</i> (-0.21)
	<i>mgs</i> (0.22)	<i>glk</i> (0.31)	<i>mgs</i> (0.25)	<i>aceAB</i> (0.42)
Vierte	ATPdrain (0.18)	<i>gnd</i> (0.26)	<i>ppsA</i> (0.16)	<i>frd</i> (0.32)
	<i>ppsA</i> (0.11)	<i>mgs</i> (0.21)	ATPdrain (0.15)	<i>edd</i> (0.30)

Ergebnisse für Lactat

Die Interventionspotentiale ausgewählter Reaktionen sind in Tabelle 6.6 gegeben. Für den Großteil der Reaktionen wird das absolute Interventionspotential mit steigendem k verstärkt. Die deutlichsten Knockout-Kandidaten unter anaeroben Bedingungen sind die Ausscheidungen von Ethanol und Acetat sowie *pfkA* (F6P::F16P), die jeweils bereits erfolgreich experimentell getestet wurden (6.1.2). Unter aeroben Bedingungen werden die ATP-Synthase (*atpA*, ATPSynth) und die Pyruvat-Dehydrogenase (*aceE*, Pyr::AcCoA) als beste Knockout-Kandidaten vorgeschlagen. Zudem werden (wie im Fall des Succinats) die in der Praxis meist verwendeten anaeroben Fermentationsbedingungen durch die Identifizierung der Sauerstoffaufnahme (O₂_up) als Knockout-Kandidat bestätigt. Der klarste und offensichtliche Überexpressionskandidat für alle Szenarien ist die Lactat-Dehydrogenase (*ldh*, bzw. Pyr::Lac). Daneben werden erneut die ATPdrain-Reaktion sowie *ppsA* (für anaerobe Bedingungen) als Überexpressionskandidaten identifiziert. Die Erklärung kann wie zuvor durch den ATP-Stoffwechsel unter Lactat-produzierenden Bedingungen gegeben werden (siehe unten).

Einfluss von CO₂, NADH und ATP auf die Lactatsynthese

Die CO₂-Aufnahmereaktion wird durch die Z₂-Werte als geringfügiger Überexpressionskandidat klassifiziert (Tabelle 6.6). Allerdings deutet der konstante Verlauf der Relevanz auf niedrigem Niveau über den gesamten Bereich von $\gamma = 0$ bis 0.9 darauf hin, dass keine gesteigerte CO₂ Supplementation notwendig ist (Abbildung 6.7). Zudem wird die CO₂-Aufnahme in der zweiten (anaerobe Szenarien) bzw. dritten Iteration (aerobe Szenarien) als Knockout-Kandidat identifiziert. Wie erwartet, ist die Bedeutung einer NADH-Quelle für alle Szenarien nur gering. Sie wird daher weder als klare Knockout- noch als Überexpressions-Intervention klassifiziert (Abbildung 6.8).

Tabelle 6.6: CASOP Z₂-Interventionspotentiale für die Produktion von Lactat. Grün hinterlegt: Übereinstimmung mit experimentell bestätigten Interventionen. Grau hinterlegt: experimentell bestätigte Intervention, kein CASOP-Kandidat. Rot hinterlegt: interessante, neu identifizierte Intervention.

Pfad	Reaktion	aerob			anaerob		
		0	2	10	0	2	10
Aufnahme	O2_up	-0,0718	-0,2386	-0,3335	-	-	-
	CO2_up	0,04	0,0732	0,1005	0,0935	0,1295	0,1033
	Glc_PTS_up	0,0098	0,0351	0,0334	0,0126	-0,0354	-0,1561
	Glc_ATP_up	0,012	0,0068	0,0061	0,0071	0,0835	0,3112
EMP	G6P::F6P	0,0648	0,1262	0,1892	0,0461	0,0237	-0,0518
	F6P::F16P/F16P::T3P	0,0625	0,1206	0,0321	-0,0091	-0,0694	-0,2102
	DHAP::G3P	-0,0127	0,0204	0,037	0,0147	-0,0643	-0,215
	G3P::DPG/DPG::3PG	0,0125	0,0412	0,0335	0	0	0
	3PG::2PG/2PG::PEP	0,0146	0,0491	0,0458	0	0	0
	PEP::PYR	0,1019	0,248	0,3623	0,0355	0,0738	-0,0022
	Pyr::PEP	0,0182	-0,028	-0,0157	0,0627	0,1489	0,2419
PPP	G6P::PGluc/PGluc::PGluc	-0,0179	-0,0357	-0,0392	-0,0039	-0,0014	0,0668
	PGluc::R15P	-0,0375	-0,1056	-0,2323	-0,0581	-0,0985	-0,0944
	R15P::X5P	-0,0058	-0,0128	-0,0321	0,012	0,0334	0,0465
	R15P::R5P	0	0	0	0	0	0
	Transket1/Transaldo	-0,0133	-0,0419	-0,0986	-0,0125	-0,0136	-0,0058
	Transket2	0,023	0,0395	0,0225	0,0309	0,0383	0,0212
EDD	2KD6PG::PyrG3P	-0,0454	-0,027	0,1563	0,025	0,093	0,3499
TCA	PYR::AcCoA	-0,0715	-0,1957	-0,3925	0,0065	0,0169	0,0996
	AcCoA::Cit	0	0	0	0	0	0
	Cit::ICit/ICit::alKG	0	0	0	0	0	0
	alKG::SuccCoA	-0,0205	-0,0623	-0,1822	-0,0631	-0,1128	-0,1236
	SuccCoA::Succ	-0,0022	0,0019	-0,0149	0,02	0,0696	0,1384
	Succ::Fum	-0,1045	-0,2606	-0,5256	-0,0703	-0,1139	-0,0817
	Fum::Succ	0,094	0,2304	0,3991	0,0732	0,1186	0,1003
	Fum::Mal	-0,0265	-0,0725	-0,1825	-0,0055	-0,0047	-0,0015
	Mal::Ox_A_NADH	-0,0055	-0,0413	-0,1533	0,0333	0,0493	0,1048
ANA	Mal::PyrNADH	0,0128	0,014	0,0936	0,0225	0,0799	0,1615
	Ox_A::PEP	0,0463	0,0189	0,0667	0,0947	0,1418	0,196
	PEP::Ox_A	0,0809	0,1813	0,329	0,0606	0,097	0,0507
GlyOx	ICit::Glyox/Glyox::Mal	-0,1134	-0,237	-0,3023	-0,1163	-0,1737	0,0047
Ferm	AcCoA::Adh/Eth_ex	-0,0236	0,1273	0,19	-0,2416	-0,4185	-0,5605
	AcCoA::AcP/AcP::Ac	0,0133	0,02	0,0327	0,0279	-0,009	-0,2482
	Pyr::Form/Form_ex	0,0251	0,1569	0,3503	-0,0903	-0,152	-0,168
	Pyr::Lac	0,1847	0,39	0,6881	0,4717	0,8071	0,963
	Pyr::Ac	0,0205	0,0747	0,1765	0,0035	0,0481	0,1138
	DHAP::Lac	0,3121	0,107	-0,0074	0,3755	0,2167	0,1973
	Lac_ex	1	1	1	1	1	1
	Ac_ex	-0,0393	-0,0697	-0,0314	-0,0582	-0,1742	-0,4005
	Succ_Ex	0,0601	0,176	0,4114	0,0391	0,0735	0,0927
ETC/AUX	NADHDehydro	-0,0296	-0,1504	-0,2284	0,0066	-0,0064	-0,0141
	Oxidase	-0,0718	-0,2386	-0,3335	0	0	0
	TransHydro	0,1649	0,2688	0,1366	-0,0028	0,0449	0,0925
	ATPSynth	-0,1263	-0,3141	-0,4241	0,0018	0,025	0,0194
	ATPdrain	0,0809	0,0823	0,1483	0,1212	0,175	0,2026
	CO2_ex	-0,1182	-0,261	-0,4144	-0,1658	-0,2485	-0,1586

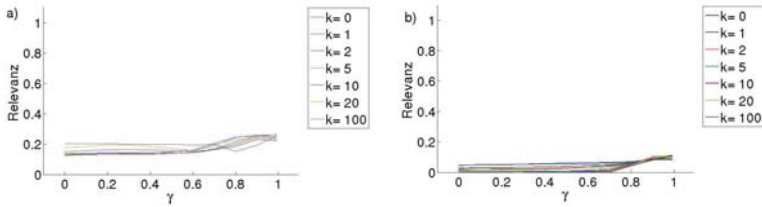


Abbildung 6.7: Relevanz der CO_2 -Aufnahme für Lactatproduktion unter a) anaeroben und b) aeroben Bedingungen für verschiedene k .

Eine ATP-Quelle kann hingegen klar als Knockout-Kandidat identifiziert werden (Abbildung 6.8). Analog wird die ATPdrain-Reaktion als Überexpressionskandidat klassifiziert. Wie im Fall von Succinat kann die Erklärung mittels der optimalen Lactatflussverteilung erfolgen. Diese ermöglicht die Generierung von bis zu zwei mol ATP pro mol NADH. Kleine γ entsprechen der Wildtypflusslage (mit hoher Biomassesynthese). Für diese wird die ATP-Aufnahme als positiv beurteilt, da sie erhöhte Biomasseausbeuten erlaubt. Wird hingegen viel Lactat gebildet, so kann dessen Ausbeute durch zusätzliches ATP nicht erhöht werden und somit wird der ATP-Aufnahmereaktion eine deutlich geringere Bedeutung beigemessen. Erneut ist diese Intervention umso deutlicher, je größer das Lactat zu Biomasse-Verhältniss ist (ab ca. 80% Lactat zu 20% Biomasse) und umso stärker der Fokus auf der Ausbeuteoptimalität liegt (Abb. 6.8).

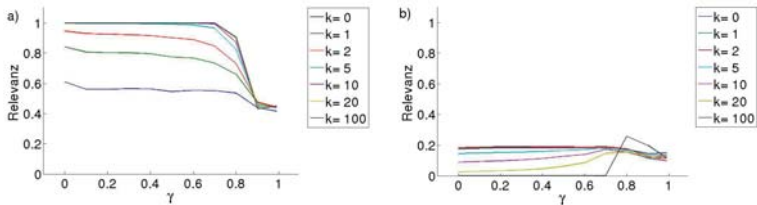


Abbildung 6.8: Relevanz einer artifiziellen a) ATP- und b) NADH-Aufnahme für Lactatproduktion unter anaeroben Bedingungen für verschiedene k .

Multiple Eingriffe zur Steigerung der Lactatsynthese

Unter anaeroben Bedingungen sind die besten Interventionskandidaten jeder Iteration identisch (Tabelle 6.7). So wird zunächst die Ethanolausscheidung (*adhE*) als bester Knockout-Kandidat der ersten Iteration identifiziert. In der zweiten Iteration wird die CO_2 -Aufnahme als bester „Knockout“ bewertet. Damit kann bereits eine Kopplung $\gamma_{\min} = 0.77$ erreicht werden. Nachdem in der dritten Iteration zudem die Pyruvat-Kinase (*pyk*) deletiert wird, erhöht sich die minimale Kopplung auf 0.83. Als bester Überexpressionskandidat wird für beide Szenarien die Pyruvat-Formiat-Lyase (*pfl*) vorgeschlagen. Auch unter aeroben Bedingungen sind die Interventionsstrategien für beide Szenarien identisch. In der ersten Iteration wird die ATP-Synthase als bester Knockout-Kandidat identifiziert, gefolgt von der Acetaldehyd-Dehydrogenase (*adhE*) in der zweiten Iteration.

Die besten Kandidaten der dritten Iteration sind die CO_2 -Aufnahmereaktion sowie die Acetat-Ausscheidung. Ohne eine Supplementierung mit CO_2 ergibt sich eine Kopplung von $\gamma_{\min} = 0.59$, die durch die zusätzliche Unterdrückung der Acetat-Ausscheidung auf 0.73 erhöht wird. Als bester Überexpressionskandidat wird der Entner-Doudoroff-Weg identifiziert (Tabelle 6.7).

Tabelle 6.7: CASOP Z_2 -Rating für die jeweils Top 3 Knockout- (Iteration 1-3) und Überexpressionskandidaten (Iteration 4). Grün hinterlegt: Übereinstimmung mit experimentell bestätigten Interventionen. Fett: in der jeweiligen Iteration ausgewählt *in silico* Eingriff.

Iteration	$S_{\text{Lac,R}}^{\text{san}}$	$S_{\text{Lac,Y}}^{\text{san}}$	$S_{\text{Lac,R}}^{\text{saer}}$	$S_{\text{Lac,Y}}^{\text{saer}}$
Erste	<i>adhE</i> (-0.24)	<i>adhE</i> (-0.42)	<i>atpA</i> (-0.13)	<i>atpA</i> (-0.31)
	<i>aceAB</i> (-0.12)	Acetat ex (-0.17)	<i>aceAB</i> (-0.11)	<i>sdh</i> (-0.26)
	<i>pfl</i> (-0.09)	<i>aceAB</i> (-0.17)	<i>sdh</i> (-0.10)	<i>aceAB</i> (-0.24)
Zweite	CO_2 up (-0.39)	CO_2 up (-0.50)	<i>adhE</i> (-0.37)	<i>adhE</i> (-0.60)
	<i>atpA</i> (-0.23)	<i>mdh</i> (-0.28)	<i>pfl</i> (-0.12)	<i>pfl</i> (-0.23)
	<i>mdh</i> (-0.22)	Acetate ex (-0.26)	<i>sdh</i> (-0.09)	Acetat ex (-0.17)
Dritte	<i>pyk</i> (-0.40)	<i>pyk</i> (-0.40)	CO_2 up (-0.80)	CO_2 up (-0.83)
	<i>pts</i> (-0.25)	<i>pts</i> (-0.25)	Acetat ex (-0.47)	Acetat ex (-0.49)
	<i>tpi</i> (-0.25)	<i>tpi</i> (-0.25)	<i>frd</i> (-0.40)	<i>frd</i> (-0.43)
Vierte	<i>pfl</i> (0.60)	<i>pfl</i> (0.62)	<i>eda</i> (0.59)	<i>eda</i> (0.60)
	<i>pta</i> (0.53)	<i>pta</i> (0.53)	<i>pts</i> (0.38)	<i>pts</i> (0.39)
	<i>mae</i> (0.50)	<i>mae</i> (0.50)	<i>pgi</i> (0.36)	<i>pgi</i> (0.37)

Ergebnisse für Lycopin

Die Interventionspotentiale ausgewählter Reaktionen sind in Tab. 6.8 gegeben. Für aerobe Bedingungen ergeben sich hier nur wenige deutliche Interventionskandidaten. Die Pyruvate-Dehydrogenase (*aceE*, $\text{PYR}::\text{AcCoA}$) sowie die ATP-Synthase (*atpA*) sind trotz ihrer kleinen absoluten Interventionspotentiale die deutlichsten Knockout-Kandidaten für ausbeute- bzw. produktivitätsorientierte Eingriffe. Weitere, schwache Interventionskandidaten, die experimentell bereits bestätigt wurden, sind der Knockout der Oxidase *cyoA* sowie die Überexpression der Triosephosphat-Isomerase (*tpi*, $\text{DHAP}::\text{G3P}$). Als triviale Überexpressionskandidaten werden die Reaktionen des DXP-Pathways sowie der Lycopin-Synthese identifiziert. Auch für anaerobe Bedingungen sind die Reaktionen des DXP-Pathways sowie der Lycopin-Synthese die klarsten Überexpressionskandidaten. Die deutlichsten Knockout-Kandidaten unter diesen Bedingungen sind die Ausscheidungen von Ethanol und Formiat. Weitere schwache (mit kleinen absoluten Interventionspotentialen) Knockout-Kandidaten sind *pykA* ($\text{PEP}::\text{Pyr}$), *sucCD* ($\text{SuccCoA}::\text{Succ}$) und *frdA* ($\text{Fum}::\text{Succ}$). Da für die anaerobe Produktion von Lycopin mit *E. coli* keine experimentellen Daten vorliegen, konnten diese Kandidaten bisher nicht validiert werden.

Einfluss von CO_2 , NADH und ATP auf die Lycopinsynthese

Die CO_2 -Aufnahme- und Ausscheidereaktion wird in Übereinstimmung mit den Analysen der optimalen Flussverteilungen entsprechend als Knockout- bzw. Überexpressionskandidat identifiziert (Tabelle 6.8). Im scheinbaren Widerspruch zur Analyse der optimalen Flussverteilung wird eine zusätzliche NADH-Quelle für alle Szenarien als unbedeutend eingestuft (Abb. 6.9).

Diese Klassifizierung kann durch die Betrachtung des ATP-Bedarfs erklärt werden. Die erreichbaren Lycopinausbeuten mit dem Stoffwechselnetz von *E. coli* betragen unter aeroben und anaeroben Bedingungen 0.09 bzw. 0.06 mol Lycopin pro mol Glukose. Die Ausbeuten werden dabei jedoch durch den ATP-Bedarf der Lycopinsynthese stärker beschränkt als durch die verfügbaren Elektronen. So ermöglicht eine externe ATP-Quelle für beide Bedingungen die zuvor bestimmte maximale Ausbeute von 0.11 mol Lycopin pro mol Glukose. Der hohe ATP-Bedarf der Lycopinsynthese wird durch die große Relevanz einer künstlichen ATP-Quelle für hohe γ bestätigt (Abb. 6.9). Analog wird die ATPdrain-Reaktion als unbedeutend eingestuft. Da dies für alle γ gilt, also auch für den Wildtyp mit Fokus auf Biomassessynthese, wird sie nicht durch das Z_2 -Rating als Knockout-Kandidat identifiziert (Tab. 6.8). Ohne externe ATP-Quelle stehen somit überschüssige Reduktionsäquivalente zur Verfügung, die mittels geeigneter Senken ausgeglichen werden müssen. Daher hat eine zusätzliche NADH-Quelle keinen positiven Einfluss und wird entsprechend niedrig bewertet.

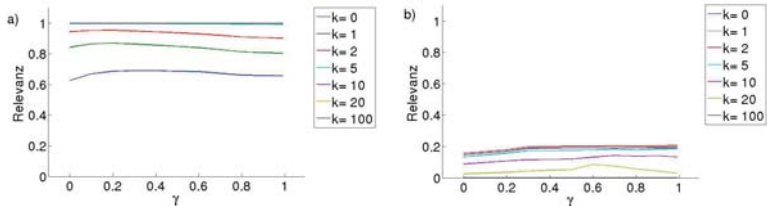


Abbildung 6.9: Relevanz einer artifiziellen a) ATP- und b) NADH-Aufnahme für Lycopinproduktion unter anaeroben Bedingungen für verschiedene k .

Multiple Eingriffe für die Lycopinsynthese

Unter anaeroben Bedingungen sind die identifizierten Strategien für beide Szenarien identisch (Tabelle 6.9). Nach der Eliminierung der Kandidaten der ersten beiden Iterationen, der Ethanol-Ausscheidung und der CO_2 -Aufnahme, wird *sucCD* als dritter Knockout-Kandidat identifiziert. Diese drei Interventionen bewirken bereits eine geringe Kopplung von $\gamma_{\min} = 0.023$. Als beste Überexpressionskandidaten werden der Glyoxylat-Weg (für $S_{\text{Succ,R}}^{\text{an}}$) bzw. die Acetat-Ausscheidung (für $S_{\text{Succ,Y}}^{\text{an}}$) identifiziert.

Unter aeroben Bedingungen unterscheiden sich die Kandidaten jeder Iteration deutlicher zwischen beiden Szenarien. Dennoch werden sowohl für $S_{\text{Succ,R}}^{\text{aer}}$ als auch für $S_{\text{Succ,Y}}^{\text{aer}}$ die Ethanolausscheidung und die ATP-Synthase als gemeinsame Knockout-Kandidaten identifiziert. Komplementiert werden die Strategien durch Knockout der Pyruvat-Dehydrogenase (*aceE*) bzw. der Pyruvat-Kinase (*pyk*) und Überexpression der Reaktionen des DXP-Pathways. Eine obligatorische Kopplung der Lycopin- an die Biomassessynthese wird unter aeroben Bedingungen mit diesen Strategien jedoch noch nicht erreicht.

Tabelle 6.8: CASOP Z₂-Interventionspotentiale für die Produktion von Lycopin. Grün hinterlegt: Übereinstimmung mit experimentell bestätigten Interventionen.

Pfad	Reaktion	aerob			anaerob		
		0	2	10	0	2	10
Aufnahme	O2_up	-0,0649	-0,0744	-0,0011	-	-	-
	CO2_up	-0,0288	-0,0114	-0,0007	-0,0972	-0,1151	-0,1553
	Glc_PTS_up	0,0356	0,0348	0,0402	0,0548	0,0449	-0,0011
	Glc_ATP_up	0,0048	-0,0206	-0,0633	0,0271	0,0662	0,0869
EMP	G6P::F6P	-0,0416	-0,0230	-0,0823	-0,0384	-0,0468	-0,0624
	F6P::F16P/F16P::T3P	-0,0205	0,0072	0,0221	-0,0201	-0,0267	-0,0517
	DHAP::G3P	0,0216	0,0504	0,1014	0,0144	-0,0237	-0,0706
	G3P::DPG/DPG::3PG	0,0006	-0,0107	-0,0363	0	0	0
	3PG::2PG/2PG::PEP	0,0019	-0,0038	-0,0253	0	0	0
	PEP::PYR	-0,0126	0,0171	-0,0273	-0,0678	-0,1311	-0,0714
	Pyr::PEP	-0,0041	-0,0151	0,0111	0,0264	0,0427	0,0385
PPP	G6P::PGluc/PGluc::PGluc	0,0175	0,0032	0,0083	0,0218	0,0212	0,0145
	PGluc::R15P	0,0405	0,0178	0,0244	0,0538	0,0706	0,1152
	R15P::X5P	-0,0103	-0,0017	0,0404	-0,0236	-0,0291	-0,0556
	R15P::R5P	0,0000	0,0000	0,0000	0	0	0
	Transket1/Transaldo	0,0094	0,0049	0,0288	0,0082	0,0119	0,0116
	Transket2	-0,0325	0,0051	0,0928	-0,0418	-0,0517	-0,0967
	2KD6PG::PyrG3P	0,0530	-0,0179	-0,1650	0,0657	0,0527	-0,0642
TCA	PYR::AcCoA	0,0228	-0,0500	-0,1051	0,1222	0,1287	0,0316
	AcCoA::Cit	0,0000	0,0000	0,0000	0	0	0
	Cit::ICit/ICit::aIKG	0,0000	0,0000	0,0000	0	0	0
	aIKG::SuccCoA	0,0172	0,0225	-0,0003	0,0233	0,0379	0,0643
	SuccCoA::Succ	-0,0151	0,0084	0,0177	-0,0402	-0,0494	-0,1117
	Succ::Fum	0,0225	-0,0244	-0,0428	0,0623	0,0579	0,0642
	Fum::Succ	-0,0184	0,0190	0,0334	-0,0823	-0,081	-0,0849
	Fum::Mal	0,0123	-0,0082	-0,0324	0,0139	0,0156	0,0423
	Mal::OxA_NADH	-0,0010	-0,0154	-0,0296	-0,0152	-0,012	0,0657
ANA	Mal::PyrNADH	0,0029	-0,0003	0,0340	-0,0017	0,004	-0,0035
	OxA::PEP	0,0297	0,0048	0,0374	0,0473	0,0456	0,0249
	PEP::OxA	-0,0249	0,0113	-0,0193	-0,0452	-0,0427	-0,066
GlyOx	ICit::Glyox/Glyox::Mal	0,0319	-0,0489	-0,0062	0,1038	0,0925	0,0183
Ferm	AcCoA::Adh/Eth_ex	-0,0497	0,0511	0,0103	-0,2904	-0,3188	-0,3387
	AcCoA::AcP/AcP::Ac	0,0029	0,0200	0,0024	0,0128	0,0547	0,0896
	Pyr::Form/Form_ex	-0,0767	0,0154	0,0581	-0,2718	-0,2764	-0,1413
	Pyr::Lac	0,0396	0,0238	0,0392	0,13	0,124	0,1193
	Pyr::Ac	0,0323	0,0312	0,0319	0,06	0,0508	-0,0128
	DHAP::Lac	-0,0437	-0,0277	0,0338	-0,0519	-0,0054	0,0446
	Lac_ex	0,0604	0,0644	0,0113	0,132	0,1356	0,1149
	Ac_ex	0,0413	0,0315	0,0121	0,0427	0,0644	0,0714
	Succ_Ex	-0,0045	0,0198	0,0394	-0,0405	-0,0349	-0,0511
	ETC/AUX	-0,0502	-0,0740	-0,0018	-0,0083	-0,0269	-0,0769
ETC/AUX	NADHDehydro	-0,0502	-0,0740	-0,0018	-0,0083	-0,0269	-0,0769
	Oxidase	-0,0649	-0,0744	-0,0011	0	0	0
	TransHydro	-0,0195	0,0234	-0,0629	0,0089	-0,002	0,0785
	ATPSynth	-0,0595	-0,1070	-0,0074	-0,021	-0,0135	0,1088
	ATPdrain	0,0179	-0,0012	-0,0005	0,0409	0,043	0,0234
DXP	CO2_ex	0,1098	0,0561	0,0042	0,2309	0,2894	0,4844
	PyrG3P::DX5P	1,0000	1,0000	1,0000	1	1	1
LycO	DMAPP::IPP	0,6667	0,6667	0,6667	0,6667	0,6667	0,6667
	IPP::GPP/GPP::FPP	1,0000	1,0000	1,0000	1	1	1
	Phytoene::Lycopene	1,0000	1,0000	1,0000	1	1	1

Tabelle 6.9: CASOP Z_2 -Rating für die jeweils Top 3 Knockout- (Iteration 1-3) und Überexpressionskandidaten (Iteration 4). Grün hinterlegt: Übereinstimmung mit experimentell bestätigten Interventionen. Fett: in der jeweiligen Iteration ausgewählt *in silico* Eingriff.

Iteration	$S_{\text{Succ,R}}^{\text{san}}$	$S_{\text{Succ,Y}}^{\text{san}}$	$S_{\text{Succ,R}}^{\text{aer}}$	$S_{\text{Succ,Y}}^{\text{aer}}$
Erste	<i>Eth ex</i> (-0.29)	<i>Eth ex</i> (-0.32)	<i>aceE</i> (-0.08)	<i>atpA</i> (-0.11)
	<i>Form ex</i> (-0.27)	<i>Form ex</i> (-0.28)	<i>atpA</i> (-0.06)	<i>ndh</i> (-0.07)
	<i>CO2 up</i> (-0.10)	<i>pyk</i> (-0.13)	<i>ndh</i> (-0.05)	<i>aceE</i> (-0.05)
Zweite	<i>CO2 up</i> (-0.57)	<i>CO2 up</i> (-0.55)	<i>Eth ex</i> (-0.08)	<i>Eth ex</i> (-0.25)
	<i>fum</i> (-0.27)	<i>fum</i> (-0.24)	<i>pgi</i> (-0.05)	<i>pyk</i> (-0.22)
	<i>sucCD</i> (-0.20)	<i>sucCD</i> (-0.22)	<i>atpA</i> (-0.05)	<i>pfl</i> (-0.13)
Dritte	<i>sucCD</i> (-0.44)	<i>sucCD</i> (-0.44)	<i>atpA</i> (-0.12)	<i>pyk</i> (-0.33)
	<i>Form ex</i> (-0.32)	<i>Form ex</i> (-0.30)	<i>ndh</i> (-0.10)	<i>Lac ex</i> (-0.23)
	<i>tkt</i> (-0.26)	<i>pyk</i> (-0.26)	<i>mgs</i> (-0.10)	<i>ndh</i> (-0.17)
Vierte	<i>aceAB</i> (0.21)	<i>Ac ex</i> (0.18)	<i>DXPpathway</i> (0.67)	<i>DXPpathway</i> (0.67)
	<i>sdh</i> (0.18)	<i>sdh</i> (0.16)	<i>pts</i> (0.24)	<i>Ac ex</i> (0.32)
	<i>Ac ex</i> (0.14)	<i>aceAB</i> (0.15)	<i>Ac ex</i> (0.20)	<i>edd/eda</i> (0.31)

Zusammenfassung der CASOP-Ergebnisse

Die mittels CASOP identifizierten Interventionsziele stimmen sehr gut mit bisherigen experimentellen Ergebnissen überein. Für alle drei untersuchten Produkte wurden viele bereits erfolgreich angewendete Kandidaten auch durch CASOP als geeignete Targets identifiziert. Darüber hinaus wurden weitere Kandidaten identifiziert, die bisher noch nicht experimentell untersucht wurden. Besonders die Überexpression von *ppsA* und (bzw. oder) der ATPdrain Reaktion für die Produktion von Succinat und Lactat stellt einen interessanten Eingriff dar, der bisher weder von anderen Methoden identifiziert noch experimentell überprüft wurde. Auch für Lycopin als Vertreter des Sekundärstoffwechsels konnten einige Kandidaten identifiziert werden, für die eine experimentelle Bestätigung vorliegt. Einige Reaktionen, die für die Lycopinsynthese bereits manipuliert wurden, sind nicht im Zentralstoffwechselmodell enthalten. Diese können trivialerweise nicht mit diesem Modell identifiziert werden.

Der Einfluss von k auf die identifizierten Interventionskandidaten führt nur sehr selten dazu, dass sich das Vorzeichen der Interventionspotentiale und damit auch die Klassifizierung als Knockout- oder Überexpressionskandidat ändert. Meist steigt das absolute Interventionspotential einer Reaktion bei steigendem k . Dies führt jedoch nur zu Änderungen der Rangfolge der Kandidaten und nicht zur Änderung der Einteilung der Interventionen in Knockout- oder Überexpressionskandidaten. Daher zeigen die multiplen Interventionsstrategien für alle hier betrachteten Produkte auch nur geringe Unterschiede zwischen Ausbeute- und produktivitätsorientierten Szenarien.

Zudem wurde CASOP bereits von anderen Arbeitsgruppen genutzt, um Interventionen zu identifizieren. So wurden zum Beispiel Knockouts zur Überproduktion von Aminosäuren mit *E. coli* mit Hilfe von CASOP bestimmt (Pande et al., 2014). In einer weiteren Studie wird ein neuer *E. coli*-Stamm konstruiert, der Succinat unter aeroben Bedingungen

mit hoher Produktivität synthetisiert (Yang et al., 2013). Dafür werden zunächst bekannte Interventionen im TCA vorgenommen (*sdh*, *iclR*) und die Ausscheidung von Nebenprodukten durch Deletionen von *poxB*, *ackA-pta* und *mgs* eliminiert. Der resultierende Stamm ZJG13 zeigt normales Wachstumsverhalten und erreicht eine spezifische Produktivität von 0.5 mmol Succinat pro Gramm Biotrockenmasse und Stunde. Anschließend wurde mit Hilfe der CASOP-Methodik die Bedeutung der drei carboxylierenden Reaktionen, Pyruvat-Carboxylase (*pyc*), Malat-Enzym (*mae*) und Phosphoenolpyruvat-Carboxylase (*ppc*) für die Succinat-Produktion analysiert. Zur Steigerung der Produktivität wird die Carboxylierung durch die Pyruvat-Carboxylase mit CASOP als die geeignetste Reaktion identifiziert. Diese Prädiktion wurde durch die Überexpression von *pyc* im Stamm ZJG13 experimentell überprüft. Der resultierende Stamm ZJG13/pT184pyc zeigt eine erhöhte spezifische Succinat-Produktivität von 0.76 mmol pro Gramm Biotrockenmasse und Stunde. Während einer Fed-Batch-Kultivierung wird eine Endkonzentration von 36.1 g/l Succinat mit einer spezifischen Produktivität von 2.75 mmol pro Gramm Biotrockenmasse und Stunde erreicht. Dies entspricht den höchsten spezifischen Produktivitäten, die bislang unter aeroben Bedingungen erreicht werden konnten.

Eine Beschränkung der CASOP-Methodik ist darin zu sehen, dass nach dem Durchlaufen der gewünschten Anzahl von Iterationen nur genau eine Strategie berechnet wurde. Wie im nächsten Abschnitt demonstriert wird, können mit den constrained Minimal Cut Sets alle möglichen Knockout-Strategien simultan berechnet werden, die ein gegebenes Strategieziel erfüllen.

6.2.2 Anwendung der constrained Minimal Cut Sets

Formulierung der Strategieziele als cMCS-Probleme

Wie zuvor werden für jedes Produkt vier verschiedene Zielstellungen definiert. So werden sowohl für aerobe (aer) als auch für anaerobe (an) Bedingungen jeweils Ausbeute- (Y) und Produktivitäts- (R) motivierte Strategien identifiziert. Jede Kombination von Produkt und Zielstellung wird einem bestimmten Szenario $\mathbb{S}_{\text{Succ/Lac/Lyco,Y/R}}^{\text{aer/an}}$ zugeordnet.

Um wohldefinierte cMCS-Probleme formulieren zu können, muss die Wahl der EM-Mengen $\mathbf{D}$ und $\mathbf{T}$ sowie des Parameters n für jedes Produkt und jede Zielstellung spezifisch angepasst werden. Als Anhaltspunkt wird für die ausbeutegerichteten Strategien gefordert, dass die Biomasseausbeute mindestens 10% der optimalen (spezifisch aeroben oder anaeroben) Ausbeute beträgt. Gleichzeitig soll die Produktausbeute mindestens 90% der optimalen (spezifisch aeroben oder anaeroben) Ausbeute betragen und mindestens ein Desired-EM soll erhalten bleiben. Für den Fall, dass diese Kriterien zu stringent sind, wird die Forderung an die Produktausbeute abgeschwächt. Als Target-EM werden alle EM gewählt, deren Produktausbeute kleiner die der Desired-EM ist.

Für die produktivitätsgerichteten Strategien wird die Anforderung an die Produktausbeute um 10% abgeschwächt. Zudem wird verlangt, dass die maximal mögliche Anzahl an Desired-EM erhalten bleibt. Dafür werden zunächst alle MCS mit $n = 1$ berechnet und anschließend die MCS selektiert, für die n maximal ist. Wie zuvor soll die Biomasseausbeute mindestens 10% der optimalen (spezifisch aeroben oder anaeroben) Ausbeute betragen.

Im Folgenden werden zunächst die jeweiligen Phänotypen anhand der EM analysiert. Anschließend werden die cMCS-Lösungen der kleinsten Kardinalitäten explizit diskutiert, da sie mit den wenigstens Interventionen zu realisieren und somit von besonderem Interesse sind. Der Kandidat „Eth_ex“ entspricht dabei einer der Reaktionen *acdH*, *adh* oder der Ethanol Ausscheidereaktion (Eth_ex). Ebenso entspricht „Lac_ex“ sowohl der Reaktion der *ldh* als auch der Lactatausscheidung. Durch „Form_ex“ werden die Reaktionen *pfl* bzw. die Formiat Ausscheidereaktion repräsentiert.

Ergebnisse für Succinat

Die vollständige Menge der EM umfasst unter aeroben und anaeroben Umweltbedingungen 225613 bzw. 49657 EM. Die konkreten Grenzwerte zur Definition der Mengen der Desired- und Target-EM sind in Tabelle 6.10 angegeben. Bei der Analyse der Phänotypenebene kann ein Bereich identifiziert werden, für den die gemeinsame Ausbeute von Succinat und Biomasse größer als 100% ist (Abbildung 6.10). Der Bereich, in dem dies möglich ist, entspricht der Fläche über der Geraden „C-mol $Y_{\text{Biomasse}} + \text{C-mol } Y_{\text{Produkt}} = 1$ “ und kann für Biomasseausbeuten bis ~ 0.05 erreicht werden. In den zugehörigen Flussverteilungen wird somit CO_2 netto fixiert. Da dieses Verhalten besonders erstrebenswert ist, werden nach den Szenarien $\mathbb{S}_{\text{Succ,Y/R}}^{\text{aer/an}}$ weitere Strategien für aerobe und anaerobe Bedingungen identifiziert, die zu obligatorischer CO_2 -Fixierung führen.

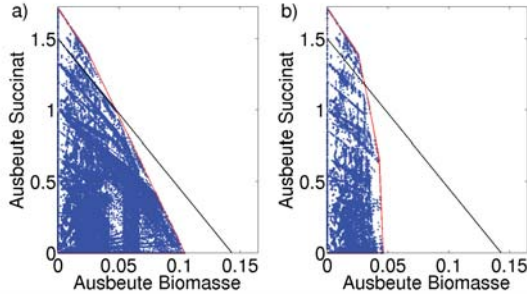


Abbildung 6.10: Phänotypebene für Biomasse- und Succinat-Ausbeute. a) aerobe Bedingungen, b) anaerobe Bedingungen. Die konvexe Hülle wird durch die roten Linien beschränkt. Die schwarze Linie entspricht einer C-mol Ausbeute von 100% (pro C-mol Glukose). Die EM werden durch die blauen Punkte repräsentiert.

Ausbeuteoptimale Strategien

Unter aeroben Bedingungen existieren mit den gegebenen Grenzwerten 59 Desired-EM sowie 225105 Target-EM (Tabelle 6.10). 2736 cMCS mit mindestens sieben und höchstens zwölf Knockouts lösen das Problem so, dass mindestens ein Desired-EM erhalten bleibt. Die zwölf cMCS, die nur sieben Knockouts benötigen, unterdrücken die Ausscheidung der Nebenprodukte Lactat, Ethanol, Acetat und Formiat. Des Weiteren müssen *aceE* und *glk* deletiert werden (Tabelle 6.10). Da zudem jedoch auch die Sauerstoffaufnahme eliminiert werden muss, sind dies keine aeroben cMCS-Lösungen im engeren Sinne. Unter anaeroben Bedingungen werden 102 EM der Menge der Desired-EM und 49195 EM der Menge der Target-EM zugeteilt. Es können 1668 zugehörige cMCS der Kardinalität fünf bis elf identifiziert werden. 18 Lösungen benötigen lediglich fünf Reaktionsdeletionen. Diese setzen sich zusammen aus der Deletion der Lactat- und Ethanolausscheidung sowie des Knockouts von *glk* oder *pgi* und der Eliminierung von entweder *ack*, *pta* oder der Acetatausscheidung (Tabelle 6.10).

Produktivitätsorientierte Strategien

Um die Zielstellung der Interventionen stärker auf die Produktivität zu legen, wird die Forderung an die Mindestproduktausbeute abgeschwächt. So muss die Succinatausbeute der Desired-EM nur 80% der optimalen Ausbeute erreichen (Tab. 6.10). Unter aeroben Bedingungen erfüllen 836 EM sowie unter anaeroben Bedingungen 903 EM die geforderten Ausbeutekriterien für Biomasse und Succinat. Die Mengen der Target-EM umfassen für Szenario $\mathbb{S}_{\text{Succ,R}}^{\text{aer}}$ 225613 EM bzw. 48352 EM für Szenario $\mathbb{S}_{\text{Succ,R}}^{\text{an}}$. Im Szenario $\mathbb{S}_{\text{Succ,R}}^{\text{aer}}$ können maximal 270 der Desired-EM erhalten bleiben. Zwölf verschiedene cMCS, die jeweils fünf Knockouts benötigen, lösen das zugehörige Problem (Tab. 6.10). Ihnen gemeinsam ist, dass sie die Ausscheidung von Lactat, Acetat, Ethanol und Formiat unterdrücken. Zudem muss auch die O_2 -Aufnahme unterdrückt werden, so dass es sich wie zuvor in Szenario $\mathbb{S}_{\text{Succ,Y}}^{\text{aer}}$ nicht um aerobe Interventionsstrategien im engeren Sinne handelt. Bei anaeroben Bedingungen können höchstens 307 der Desired-EM erhalten bleiben. Sechs cMCS der Kardinalität vier lösen das zugehörige Problem (Tab. 6.10). Diese sechs Strategien unterdrücken die Lactat-, Ethanol-, Formiat- und Acetatausscheidung.

Tabelle 6.10: cMCS-Problemd Definitionen und Lösungen für Succinat. $Y^{X/S}$: Ausbeute g Biomasse / mol Glukose, $Y^{S_{succ}}$: Ausbeute mol Succinat / mol Glukose.

Szenario	D	T	n	# cMCS	Kardinalitäten	Kleinste Lösungen
$\mathbb{S}_{Succ,Y}^{Succ}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.01 \wedge Y_E^{S_{succ}} \geq 1.54\}$	$\{E \in \mathbf{E} Y_E^{S_{succ}} < 1.54\}$	1	2736	7-12	$Ac_ex \wedge Lac_ex \wedge glk \wedge ace \wedge O_2_up \wedge Form_ex \wedge Eth_ex$
$\mathbb{S}_{Succ,Y}^{San}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.0046 \wedge Y_E^{S_{succ}} \geq 1.54\}$	$\{E \in \mathbf{E} Y_E^{S_{succ}} < 1.54\}$	1	1668	5-11	$glk \wedge pgi \wedge Lac_ex \wedge Eth_ex \wedge \{pta \vee ack \vee Ac_ex\}$
$\mathbb{S}_{Succ,R}^{Succ}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.01 \wedge Y_E^{S_{succ}} \geq 1.37\}$	$\{E \in \mathbf{E} Y_E^{S_{succ}} < 1.37\}$	$max!$	12	5	$Ac_ex \wedge Lac_ex \wedge Form_ex \wedge Eth_ex$
$\mathbb{S}_{Succ,R}^{San}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.0046 \wedge Y_E^{S_{succ}} \geq 1.37\}$	$\{E \in \mathbf{E} Y_E^{S_{succ}} < 1.37\}$	$max!$	6	4	O_2_up $Ac_ex \wedge Lac_ex \wedge Form_ex \wedge Eth_ex$
$\mathbb{S}_{Succ,CO_2,Y}^{Succ}$	$\{E \in \mathbf{E} Y_E^{X/S} > 0 \wedge (Y_E^{X/S} + Y_E^{S_{succ}}) > 100\%\}$	$\{E \in \mathbf{E} (Y_E^{X/S} + Y_E^{S_{succ}}) \leq 100\%\}$	1	2366	4-11 und 13	$Ac_ex \wedge Lac_ex \wedge Eth_ex \wedge O_2_up$
$\mathbb{S}_{Succ,CO_2,Y}^{San}$	$\{E \in \mathbf{E} Y_E^{X/S} > 0 \wedge (Y_E^{X/S} + Y_E^{S_{succ}}) > 100\%\}$	$\{E \in \mathbf{E} (Y_E^{X/S} + Y_E^{S_{succ}}) \leq 100\%\}$	1	1285	3-12	$Ac_ex \wedge Lac_ex \wedge Eth_ex$
$\mathbb{S}_{Succ,CO_2,R}^{Succ}$	$\{E \in \mathbf{E} Y_E^{X/S} > 0 \wedge (Y_E^{X/S} + Y_E^{S_{succ}}) > 100\%\}$	$\{E \in \mathbf{E} (Y_E^{X/S} + Y_E^{S_{succ}}) \leq 100\%\}$	$max!$	6	5	$Ac_ex \wedge Lac_ex \wedge Eth_ex \wedge O_2_up$
$\mathbb{S}_{Succ,CO_2,R}^{San}$	$\{E \in \mathbf{E} Y_E^{X/S} > 0 \wedge (Y_E^{X/S} + Y_E^{S_{succ}}) > 100\%\}$	$\{E \in \mathbf{E} (Y_E^{X/S} + Y_E^{S_{succ}}) \leq 100\%\}$	$max!$	3	3	$Ac_ex \wedge Lac_ex \wedge Eth_ex$

Die Ausbeute der ersten vier Strategien wird ist mol Succinat pro mol Glukose, die Ausbeute der letzten vier Strategien wird als „(C-mol Biomasse plus C-mol Succinat) pro C-mol Glukose“ berechnet.

Strategien zur Fixierung von CO₂

Elementarmoden, bei denen CO₂ fixiert werden kann, sind von besonderem Interesse. Sie ermöglichen eine gemeinsame Ausbeute von mehr als 1 C-mol Biomasse und Succinat pro C-mol Glukose ($Y^{u+Succ} > 1$). Um zu garantieren, dass ausschließlich CO₂-fixierende Stoffwechselwege funktional sind, werden die cMCS-Probleme wie folgt aufgestellt: In die Menge der Desired-EM werden alle EM selektiert, die sowohl eine positive Wachstumsrate haben und deren gemeinsame Ausbeute > 1 ist. Die CO₂-Aufnahmereaktion ist in allen diesen EM aktiv. Die Menge der Target-EM setzt sich aus den EM zusammen, deren gemeinsame Succinat- und Biomasseausbeute ≤ 1 ist.

Analog zu der Identifizierung der vorherigen Strategien werden zwei Szenarien für aerobe und anaerobe Bedingungen betrachtet. In den Szenarien $\mathbb{S}_{Succ,CO_2,Y}^{aer/an}$ genügt es, mindestens einen EM aus **D** zu erhalten. Für die Szenarien $\mathbb{S}_{Succ,CO_2,R}^{aer/an}$ soll analog zu den Produktivitätsstrategien die maximale Anzahl an Desired-EM erhalten bleiben (Tabelle 6.10). Die Menge der Desired-EM umfasst 1233 EM für die aeroben Szenarien $\mathbb{S}_{Succ,CO_2,Y/R}^{aer}$ sowie 1195 EM für die anaeroben Szenarien $\mathbb{S}_{Succ,CO_2,Y/R}^{an}$. Entsprechend beinhalten die Target-EM Mengen 223084 und 48239 EM für die Szenarien $\mathbb{S}_{Succ,CO_2,Y/R}^{aer}$ bzw. $\mathbb{S}_{Succ,CO_2,Y/R}^{an}$.

Im aeroben Fall existieren für Szenario $\mathbb{S}_{Succ,CO_2,Y}^{aer}$ 2366 cMCS-Lösungen mit Kardinalitäten zwischen vier und 13. Die sechs cMCS der kleinsten Kardinalität umfassen die Eliminierung der Ausscheidung von Acetat, Lactat und Ethanol sowie die Deletion der Sauerstoffaufnahme. Dies sind demnach keine aeroben Interventionen im engeren Sinne. Von den 2366 cMCS-Lösungen sind 684 „echte“ aerobe Interventionsstrategien, da diese nicht die Eliminierung der Sauerstoffaufnahme benötigen. Sie benötigen zwischen fünf und elf Knockouts, wobei die zwölf kleinsten Lösungen (mit fünf bzw. sechs Knockouts) experimentell nicht realisierbar sind, da sie die CO₂-Ausscheidung als Knockout-Kandidaten beinhalten. Die verbleibenden cMCS umfassen mindestens acht Deletionen, auf dessen Darstellung hier verzichtet wird. Sechs cMCS der Größe vier lösen das Problem von Szenario $\mathbb{S}_{Succ,CO_2,R}^{aer}$, so dass die maximale Anzahl an Desired-EM erhalten bleiben (635 EM). Diese cMCS entsprechen denen des Szenarios $\mathbb{S}_{Succ,CO_2,Y}^{aer}$ und sind daher unter aeroben Bedingungen nicht zugänglich (Tabelle 6.10).

Unter anaeroben Bedingungen lösen 1285 cMCS mit Kardinalitäten von drei bis zwölf das Problem des Szenario $\mathbb{S}_{Succ,CO_2,Y}^{an}$. Die drei Lösungen der Größe drei unterdrücken die Ausscheidung von Acetat, Lactat und Ethanol. Das Problem des Szenarios $\mathbb{S}_{Succ,CO_2,R}^{an}$ wird von den selben drei cMCS gelöst. Es bleiben erneut 635 Desired-EM erhalten.

Ergebnisse für Lactat

Die optimale (addierte) Ausbeute ist hier auf 100% der eingesetzten Glukose beschränkt und wird nur dann erreicht, wenn keine Biomasse gebildet wird. Sobald ein Teil des Kohlenstoffs für die Synthese der Biomasse genutzt wird, muss CO_2 ausgeschieden werden, das bei der Laktatsynthese nicht refixiert werden kann. Wie zuvor beinhaltet die vollständige Menge der EM unter aeroben und anaeroben Umweltbedingungen 225613 bzw. 49657 EM.

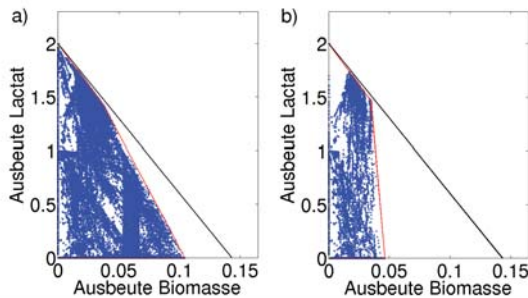


Abbildung 6.11: Phänotypenebene für Biomasse- und Lactat-Ausbeute. a) aerobe Bedingungen, b) anaerobe Bedingungen. Die konvexe Hülle wird durch die roten Linien beschränkt. Die schwarze Linie entspricht einer C-mol Ausbeute von 100% (pro C-mol Glukose). Die EM werden durch die blauen Punkte repräsentiert.

Ausbeuteoptimale Strategien

Unter aeroben Bedingungen erfüllen 29 EM die Anforderungen der Desired-EM. Der Menge der Target-EM werden 224209 EM zugeteilt. Es existieren 66 cMCS mit mindestens sieben und höchstens neun Deletionen, so dass mindestens ein EM der Menge der Desired-EM nicht betroffen ist. Von diesen benötigen zwölf cMCS nur sieben Knockouts. Sie unterdrücken jeweils die Ausscheidung von Ethanol und Formiat. Darüber hinaus müssen *atpA*, *pntA*, *ppc*, *poxB* sowie *sucCD* oder *tkt* deletiert werden (Tabelle 6.11). Die resultierenden Knockout-Strategien sind sehr komplex und daher experimentell vermutlich schwer zu implementieren.

Unter anaeroben Bedingungen gibt es keinen EM, der die Forderung nach einer Biomasseausbeute von 10% mit gleichzeitiger Produktausbeute von 90% (1.8 mol Lactat / mol Glukose) des jeweiligen Optimums erfüllt. Daher wird die Forderung der Produktausbeute auf 1.7 mol Lactat / mol Glukose abgeschwächt (Tabelle 6.11). Die Menge der Desired- und Target-EM beinhalten dann entsprechend 331 bzw. 49299 EM. 618 cMCS mit Kardinalitäten von vier bis zehn lösen dieses Problem. Die 18 cMCS der Kardinalität vier deletieren die Ethanolausscheidung und *pts* sowie jeweils eine Reaktion der oberen Glykolyse und des TCA (Tabelle 6.11).

Tabelle 6.11: cMCS-Problemlösungen und Lösungen für Lactat. $Y^{X/S}$: Ausbeute g Biomasse / mol Glukose, Y^{Lac} : Ausbeute mol Lactat / mol Glukose.

Szenario	D	T	n	# cMCS	Kardinalitäten	Kleinste Lösungen
$\mathbb{S}_{Lac,Y}^{Sacr}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.01 \wedge Y_E^{Lac} \geq 1.8\}$	$\{E \in \mathbf{E} Y_E^{Lac} < 1.8\}$	1	66	7-9	$ppc \wedge poxB \wedge$ $atpA \wedge pntA \wedge$ $Form_ex \wedge Eth_ex \wedge$ $\{sucCD \vee tkt\}$
$\mathbb{S}_{Lac,Y}^{San}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.0046 \wedge Y_E^{Lac} \geq 1.7\}$	$\{E \in \mathbf{E} Y_E^{Lac} < 1.7\}$	1	618	4-10	$pts \wedge Eth_ex \wedge$ $\{frd \vee succ_ex\} \wedge$ $\{pfk \vee fba \vee tpi\}$
$\mathbb{S}_{Lac,R}^{Sacr}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.01 \wedge Y_E^{Lac} \geq 1.6\}$	$\{E \in \mathbf{E} Y_E^{Lac} < 1.6\}$	$max!$	6	5	$CO_2_up \wedge Ac_ex \wedge$ $atpA \wedge Eth_ex$ $\{pfk \vee fba\}$
$\mathbb{S}_{Lac,R}^{San}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.0046 \wedge Y_E^{Lac} \geq 1.6\}$	$\{E \in \mathbf{E} Y_E^{Lac} < 1.6\}$	$max!$	9	4	$CO_2_up \wedge aceE \wedge Eth_ex \wedge$ $\{pfk \vee fba \vee tpi\}$

Produktivitätsorientierte Strategien

Unter aeroben Bedingungen werden 8767 EM als Desired-EM selektiert. Diese synthetisieren Biomasse und Lactat mit einer Ausbeute von mindestens 10 bzw. 80% der (spezifisch aeroben) optimalen Ausbeute. 3075 EM erfüllen diese Anforderungen für anaerobe Bedingungen und werden daher als Desired-EM selektiert. Die Mengen der Target-EM umfassen 214593 und 46547 EM unter aeroben bzw. anaeroben Umweltbedingungen.

Von den aeroben Desired-EM können maximal 540 erhalten werden. Sechs verschiedene cMCS mit jeweils fünf Knockouts existieren, um dieses Problem zu lösen. Diese fünf Lösungen enthalten jeweils die CO_2 -Aufnahme, so dass lediglich vier Knockouts durchgeführt werden müssen. Diese umfassen *adhE*, *atpA*, die Acetatausscheidung sowie *pfk* oder *fba*. Unter anaeroben Bedingungen können bis zu 312 der Desired-EM erhalten bleiben. Dies ist mit neun cMCS der Kardinalität vier möglich, wobei auch diese die Deletion der CO_2 -Aufnahme erfordern, so dass lediglich drei Knockouts (*aceE*, *adhE* sowie eine Reaktion der oberen Glykolyse) implementiert werden müssen (Tabelle 6.11).

Ergebnisse für Lycopin

Die optimale (addierte) Ausbeute ist hier deutlich kleiner als 100% der eingesetzten Glukose (Abbildung 6.12). Zudem ergeben sich unter anaeroben Bedingungen im Vergleich zu aeroben Bedingungen geringere optimale Produktausbeuten. Die vollständige Menge der Elementarmoden unter aeroben und anaeroben Umweltbedingungen beinhaltet 353196 bzw. 104103 EM.

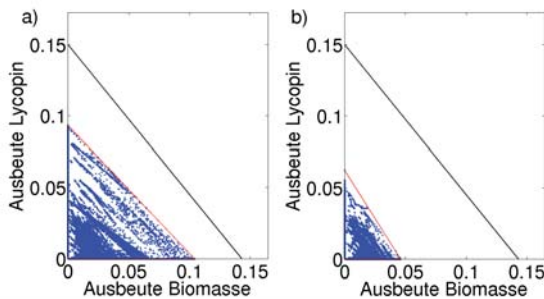


Abbildung 6.12: Phänotypebene für Biomasse- und Lycopin-Ausbeute. a) aerobe Bedingungen, b) anaerobe Bedingungen. Die konvexe Hülle wird durch die roten Linien beschränkt. Die schwarze Linie entspricht einer C-mol Ausbeute von 100% (pro C-mol Glukose). Die EM werden durch die blauen Punkte repräsentiert.

Ausbeuteoptimale Strategien

Für beide Szenarien, $\mathbb{S}_{\text{LycO},Y}^{\text{aer}}$ und $\mathbb{S}_{\text{LycO},Y}^{\text{an}}$, existieren keine EM, die die Bedingungen an die Biomasse- und Lycopinausbeute erfüllen. Unter aeroben Bedingungen wird die geforderte Mindest-Lycopinausbeute daher auf 80% abgeschwächt. Mit diesem angepassten Schwellenwert beinhaltet die Menge der Desired-EM 27 EM. Unter anaeroben Bedingungen ist eine Herabsetzung auf 70% der optimalen Lycopinausbeute notwendig, so dass die Menge der Desired-EM 33 EM umfasst (Tabelle 6.12). Die Target-Mengen beinhalten 349063 EM unter aeroben sowie 103603 EM unter anaeroben Bedingungen.

Für das Szenario $\mathbb{S}_{\text{LycO},Y}^{\text{aer}}$ werden 62 cMCS identifiziert, die zwischen sechs und elf Reaktionsdeletionen benötigen. Es existieren zwei cMCS-Lösungen, die lediglich sechs Knockouts benötigen (Tabelle 6.12). Diese umfassen *pntA*, *mae*, *aceE*, *pck*, *pgi* sowie eine der Reaktionen des Entner-Doudoroff-Weges *edd* oder *eda*.

Im anaeroben Szenario $\mathbb{S}_{\text{LycO},Y}^{\text{an}}$ können zwölf cMCS der Kardinalität sechs identifiziert werden. Diese setzen sich zusammen aus den Deletionen von *glk*, *poxB*, *atpA* sowie der Unterbindung der Ausscheidung von Lactat, Ethanol und Formiat (Tabelle 6.12).

Produktivitätsorientierte Strategien

Wie zuvor wird in den Szenarien $\mathbb{S}_{\text{LycO},R}^{\text{aer}}$ und $\mathbb{S}_{\text{LycO},R}^{\text{an}}$ die Anforderung an die Produktausbeute leicht abgeschwächt (Tabelle 6.12). Unter aeroben Bedingungen genügt hier eine Lycopinausbeute von 0.065 (70% des aeroben Maximums). Entsprechend genügt bei anaeroben Bedingungen eine Lycopinausbeute von 0.038 (60% des anaeroben Maximums). Mit diesen Grenzwerten werden 435 (in $\mathbb{S}_{\text{LycO},R}^{\text{aer}}$) bzw. 495 EM (in $\mathbb{S}_{\text{LycO},R}^{\text{an}}$) den Desired-EM zugewiesen. Die Mengen der Target-EM umfassen entsprechend 346813 ($\mathbb{S}_{\text{LycO},R}^{\text{aer}}$) bzw. 102166 EM ($\mathbb{S}_{\text{LycO},R}^{\text{an}}$).

Unter aeroben Bedingungen können höchstens 51 der Desired-EM erhalten bleiben. Zehn cMCS-Lösungen mit sechs oder sieben Knockouts lösen das entsprechende Problem. Die Interventionsstrategien mit sechs Deletionen setzen sich zusammen aus den Knockouts von *ppc*, *mgs*, *pyk*, einer der beiden Reaktionen des Entner-Doudoroff-Weges (*edd* oder *eda*) sowie zwei Reaktionen aus der Menge von $\{pgi, aceE, pntA\}$. Für Szenario $\mathbb{S}_{\text{LycO},R}^{\text{an}}$ existieren zwölf cMCS der Größe fünf, so dass 78 Desired-EM erhalten bleiben. Diese cMCS setzen sich zusammen aus den Knockouts von *glk* und *ppsA* sowie der Eliminierung der Ausscheidung der Nebenprodukte Lactat, Formiat und Ethanol (Tabelle 6.12).

Tabelle 6.12: cMCS-Problemlösungen und Lösungen für Lycopin. $Y^{X/S}$: Ausbeute g Biomasse / mol Glukose, Y^{Lyc} : Ausbeute mol Lycopin / mol Glukose.

Szenario	D	T	n	# cMCS	Kardinalitäten	Kleinste Lösungen
$S_{Lyc,Y}^{sacr}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.01 \wedge Y_E^{Lyc} \geq 0.075\}$	$\{E \in \mathbf{E} Y_E^{Lyc} < 0.075\}$	1	62	6,7,10,11	$pntA \wedge mae \wedge aceE \wedge$ $pck \wedge pgi \wedge$ $\{edd \vee eda\}$
$S_{Lyc,Y}^{san}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.0046 \wedge Y_E^{Lyc} \geq 0.044\}$	$\{E \in \mathbf{E} Y_E^{Lyc} < 0.044\}$	1	12	6	$glk \wedge atpA \wedge$ $poxB \wedge Lac_{ex} \wedge$ $Form_{ex} \wedge Eth_{ex}$
$S_{Lyc,R}^{sacr}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.01 \wedge Y_E^{Lyc} \geq 0.065\}$	$\{E \in \mathbf{E} Y_E^{Lyc} < 0.065\}$	$max!$	10	6,7	$ppc \wedge mgs \wedge pyk \wedge$ $\{edd \vee eda\} \wedge$ $\{pgi \wedge aceE\} \vee$ $\{pgi \wedge pntA\} \vee$ $\{aceE \wedge pntA\}$
$S_{Lyc,R}^{san}$	$\{E \in \mathbf{E} Y_E^{X/S} \geq 0.0046 \wedge Y_E^{Lyc} \geq 0.038\}$	$\{E \in \mathbf{E} Y_E^{Lyc} < 0.038\}$	$max!$	12	5	$glk \wedge ppsA \wedge Lac_{ex} \wedge$ $Form_{ex} \wedge Eth_{ex}$

Zusammenfassung der cMCS-Ergebnisse

Die mittels der cMCS identifizierten Knockout-Strategien für die Produktion von Succinat stimmen sehr gut mit experimentellen Ergebnissen überein. So wurden für alle Szenarien die Ausscheidungen der Nebenprodukte Ethanol, Lactat, Acetat und Formiat als Knockout-Kandidaten identifiziert. Ebenso wurde für alle aeroben Szenarien die Sauerstoffaufnahme als Knockout-Kandidat bestimmt, womit die Vorteile anaerober Fermentationsbedingungen bestätigt wurden. Eine interessante neue Knockout-Strategie wurde für Szenario $\mathbb{S}_{\text{Succ},Y}^{\text{an}}$ identifiziert. Die Strategien mit nur fünf Deletionen eliminieren nicht das (experimentell bestätigte und auch von CASOP vorgeschlagene) Zucker-PTS-System (*ptsA*), sondern die alternative Glukoseaufnahme via Glukokinase (*glk*). Die Kombination mit dem Knockout der Phosphoglukose Isomerase (*pgi*) bewirkt, dass der Entner-Doudoroff-Weg essentiell wird und über diesen zusätzliches (zu dem über das Zucker-PTS-System gebildetem) Pyruvat gebildet wird. Da zudem die Ausscheidung von Ethanol, Lactat und Acetat unterdrückt ist, muss dieses über die (nun ebenfalls essentielle) Phosphoenolpyruvat Synthase (*ppsA*) zum Succinat katabolisiert werden, um die Redox-äquivalente auszugleichen. Ein weiteres interessantes Resultat ist, dass für eine obligatorische CO_2 -Fixierung mit *E. coli* unter anaeroben Bedingungen bereits drei Knockouts genügen. Diese Knockouts umfassen die Eliminierung der Ausscheidung von Ethanol, Lactat und Acetat. Die Deletion dieser drei Reaktionen erzwingt zudem die CO_2 -Fixierung mit den wenigsten Nebeneffekten (es bleiben viele EM im Netzwerk funktional).

Für die Produktion von Lactat konnten fast alle experimentell validierte Knockouts bestätigt werden. So wurden z.B. die Ausscheidung von Ethanol, Acetat und Formiat, die Pyruvat-Dehydrogenase (*aceE*), die ATP-Synthase (*atpA*) sowie die Phosphofruktokinase *pfkA* und die Fruktosebisdiphosphataldolase *fba* bestätigt. Zudem konnten auch konkrete Kombinationen dieser Knockouts in experimentellen Studien wiedergefunden werden. Dass CO_2 -Supplementierung für die Lactatproduktion keine Bedeutung hat, wurde durch deren Bestimmung als „Knockout-Kandidat“ bestätigt.

Die experimentelle Übereinstimmung der vorgeschlagenen Knockout-Kandidaten für die Produktion von Lycopin ist deutlich geringer. Zwar wurden auch hier einige Kandidaten wie z.B. *aceE*, *ppc*, *mae* und *poxB* bestätigt, allerdings konnten für keine der vollständigen vorgeschlagenen Kombinationen experimentelle Bestätigungen gefunden werden. Da jedoch noch keine Mutanten mit den hier angestrebten Lycopinausbeuten existieren, kann daraus nicht geschlossen werden, dass die vorgeschlagenen Knockout-Strategien nicht zutreffend sind. Vielmehr können sie als Ausgangspunkte für zukünftige experimentelle Überprüfungen dienen. In (Gruchattka et al., 2013) wurden für eine vergleichende Analyse des Potentials zur Produktion des Lycopin-Präkursors Isopentenyl Diphosphat (IPP) in *E. coli* und *S. saccharomyces* Interventionskandidaten identifiziert. Der Ansatz der cMCS wurde in dieser Studie für die Bestimmung von Knockout-Strategien zur IPP-Produktion genutzt. Dabei wurde ein anderes Stoffwechselmodell sowie andere Grenzwerte für die Definition der Desired-EM und Target-EM verwendet, so dass andere cMCS als die hier präsentierten berechnet wurden. Einer der in (Gruchattka et al., 2013) zur obligatorischen (aeroben) Lycopinproduktion identifizierten cMCS benötigt nur zwei Reaktionsdeletionen. So führt die Deletion der Transhydrogenase *pntA* sowie der Enolase *eno* dort zu einer Mindestausschüttung von 0.34 mol IPP / mol Glukose bzw 0.04 mol Lycopin / mol Glukose. Durch den Knockout der Enolase muss in dem verwendeten Netzwerk

der Entner-Doudoroff-Pathway benutzt werden, was zu einem NADPH-Überschuss führt. Dieser kann nach der Deletion der Transhydrogenase nur über den DXP-Weg reoxidiert werden und führt somit zu einer obligatorischen Lycopin-Ausscheidung. In der vorliegenden Arbeit wurde dieser cMCS nicht identifiziert, da das hier verwendete Netzwerk zusätzlich den Methylglyoxylat-Weg (über *mgs*) umfasst. Über diesen kann Glukose katabolisiert werden, ohne NADPH im Überschuss zu generieren. Die hier präsentierten aeroben Knockout-Strategien verfolgen dennoch dasselbe zu Grunde liegende Prinzip. So werden die Flüsse derart umgelenkt, dass ein Überschuss an NADPH erzeugt wird, der nach der Deletion der Transhydrogenase nur über die Lycopinsynthese regeneriert werden kann. Dieses Beispiel demonstriert zugleich eine Limitierung von Modellen des Zentralstoffwechsels für die Identifizierung von Eingriffen für Produkte, deren Synthese außerhalb der zentralen Pathways stattfindet. So können identifizierte Interventionen sensitiv gegenüber den im Modell berücksichtigten Pathways sein, selbst wenn diese nicht unmittelbar mit der Produktsynthese verknüpft sind. Eine Verzerrung der Ergebnisse kann durch die Verwendung genomskaliger Netzwerke vermieden werden, da diese das vollständige Potential aller metabolischen Reaktionen eines Organismus umfassen. Im nächsten Abschnitt wird daher eine Anwendung des EFMSampler demonstriert, der elementarmodenbasierte Methoden auch für diese Netzwerke zugänglich macht.

6.2.3 Anwendung des EFMSampler

In Kapitel 5.3 wurde gezeigt, dass der EFMSampler weitgehend unverzerrte Elementarmodenteilmengen generiert. Diese können zur Approximation der Eigenschaften der vollständigen Menge der EM benutzt werden. Da die vollständigen Elementarmodenteilmengen für genomskalige Netzwerke nicht berechnet werden können, sind diese im Allgemeinen nicht für die Analyse durch EM-basierte Methoden zugänglich. Im Folgenden wird der EFMSampler verwendet, um mit der CASOP-Methodik Interventionskandidaten zur Produktion von Succinat, Lactat und Lycopin im genomskaligen *E. coli*-Stoffwechselnetz iAF1260 zu identifizieren. Die Parameter der CASOP-Methode werden analog zu den Fallstudien im Zentralstoffwechsel-Modell (ZSW-Modell) gewählt ($\gamma = 0/\gamma = 0.9$, $k = 0/k = 2$ und Bewertungsschema Z_2). Mit Hilfe des EFMSamplers werden für jedes Produkt vier verschiedene EM-Mengen (aerobe und anaerobe Bedingungen mit $\gamma = 0$ bzw. 0.9) mit jeweils 250.000 EM generiert (insgesamt 1 Million EM je Produkt). Die Demonstration der Anwendung des EFMSampler steht dabei im Vordergrund, so dass nur die Knockout- und Überexpressionskandidaten der ersten Iteration bestimmt werden. Zudem wird die Betrachtung hier auf die jeweils relevanten Produktionsbedingungen, das heißt anaerobe Bedingungen für die Produktion von Succinat und Lactat sowie aerobe Bedingungen für die Synthese von Lycopin beschränkt.

Übereinstimmung mit Kandidaten des Zentralstoffwechselmodells und experimentell validierten Kandidaten

Für Succinat sind von den zehn präsentierten Knockout-Kandidaten des GS-Modells sechs auch mit dem ZSW-Modell identifiziert worden (Tab. 6.13). In Übereinstimmung mit experimentell validierten Kandidaten konnte die Ethanol-Ausscheidung und der Bedarf an CO₂ korrekt identifiziert werden. Die weiteren offensichtlichen Knockout-Kandidaten Acetat- und Format-Ausscheidung gehören zwar nicht zu den besten Kandidaten, werden jedoch auch als schwache Knockout-Kandidaten identifiziert. Von den zehn besten Überexpressionskandidaten stimmen vier mit denen des ZSW-Modells überein. Als einer der besten Kandidaten wurde die experimentell bestätigte Überexpression des Glyoxylat-Wegs identifiziert.

Es sei daran erinnert, dass der EFMSampler reversible Reaktionen vor Berechnung der Elementarmoden in die Hin- und Rückrichtung aufsplittet. Somit können die Relevanzen dieser Reaktionen für beide Richtungen getrennt bestimmt werden. Dies erklärt auch, dass die reversible Fumarase (*fum*) hier sowohl als Knockout- als auch als Überexpressionskandidat identifiziert wurde, je nachdem, ob sie in oxidativer oder reduktiver Richtung abläuft. Dieselbe Argumentation gilt für die Malat Dehydrogenase (*mdh*). Die integrierte Relevanz der Hin- und Rückrichtung aller reversiblen Reaktionen lässt sich aus der Summe der Relevanzen der jeweiligen Hin- und Rückrichtung berechnen. Im vorliegenden Fall würden sowohl die reversible Fumarase als auch die reversible Malat Dehydrogenase in Übereinstimmung mit experimentellen Daten nicht als Interventionskandidat in Erscheinung treten, da die Relevanzen der Hin- und Rückrichtung in etwa gleich groß sind. Dennoch kann die Information über die vorteilhafte Richtung einer reversiblen Reaktion von zusätzlichem Nutzen sein.

Für Lactat stimmen jeweils vier der besten zehn Knockout- und Überexpressionskandidaten des GS-Modells mit denen des ZSW-Modells überein (Tab. 6.13). Die Übereinstimmung mit experimentell validierten Knockouts ist hier besonders gut. So werden im Gegensatz zum ZSW-Modell die ATP Synthase (*atpA*) und die Fumarat Reduktase (*frd*) bereits in der ersten Iteration identifiziert. Für die Lactatsynthese mit *E. coli* wurden bisher keine Überexpressionskandidaten experimentell validiert, so dass die hier präsentierten Kandidaten als Ausgangspunkt für spätere experimentelle Überprüfungen dienen können.

Für Lycopin ist die geringste Übereinstimmung der Knockout-Kandidaten zu den Ergebnissen des ZSW-Modells zu beobachten. Lediglich die ATP-Synthase (*atpA*) sowie die Oxidase (*cyoA*) wurden auch mit dem ZSW-Modell identifiziert. Hier konnten zusätzlich die experimentell bestätigten Knockouts der Fumarat Reduktase (*frd*) sowie der Format Hydrogenylase (*fdhF*) identifiziert werden. Die Pyruvat Dehydrogenase (*aceE*) gehört zwar nicht zu den besten Kandidaten, wurde jedoch in Übereinstimmung mit experimentellen Daten ebenfalls als Knockout-Kandidat identifiziert. Als offensichtliche Überexpressionskandidaten wurden, wie auch mit dem ZSW-Modell, die Reaktionen des DXP-Pathways und der Lycopinsynthese identifiziert. Diese Ergebnisse bestätigen, dass der EFMSampler geeignet ist, elementarmodenbasierte Verfahren in genomskaligen Netzwerken zur Anwendung zu bringen.

Tabelle 6.13: CASOP Z₂-Rating der Top 5 Knockout- und Überexpressionskandidaten der ersten Iteration im genomskaligen Netzwerk (iAF1260). Fettgedruckt: Reaktionen, die auch im Zentralstoffwechselmodell entsprechend identifiziert worden sind. Grün hinterlegt: Übereinstimmung mit experimentell bestätigten Interventionen.

Succinat				
	Knockout-Kandidaten		Überexpressionskandidaten	
	$\mathbb{S}_{\text{Succ,R}}^{\text{an}}$	$\mathbb{S}_{\text{Succ,Y}}^{\text{an}}$	$\mathbb{S}_{\text{Succ,R}}^{\text{an}}$	$\mathbb{S}_{\text{Succ,Y}}^{\text{an}}$
1	<i>fum</i> (-0.51)	<i>fum</i> (-0.64)	<i>fum</i> -back (0.50)	<i>fum</i> -back (0.72)
2	<i>mdh</i> (-0.34)	<i>mdh</i> (-0.44)	<i>mdh</i> -back (0.34)	<i>mdh</i> -back (0.58)
3	Glycolate_out (-0.21)	CO₂_ex (-0.33)	aceB (0.30)	<i>nuoA</i> (0.36)
4	<i>sucCD</i> (-0.13)	<i>pykA</i> (-0.22)	aceA (0.29)	<i>glkA</i> (0.31)
5	CO₂_ex (-0.07)	<i>adhE</i> (-0.20)	<i>nuoA</i> (0.27)	aceA (0.27)
Lactat				
	Knockout-Kandidaten		Überexpressionskandidaten	
	$\mathbb{S}_{\text{Lac,R}}^{\text{an}}$	$\mathbb{S}_{\text{Lac,Y}}^{\text{an}}$	$\mathbb{S}_{\text{Lac,R}}^{\text{an}}$	$\mathbb{S}_{\text{Lac,Y}}^{\text{an}}$
1	<i>frd</i> (-0.11)	<i>atpA</i> (-0.27)	<i>ldh</i> (0.75)	<i>ldh</i> (0.97)
2	<i>atpA</i> (-0.09)	<i>adhE</i> (-0.26)	<i>hchA</i> (0.12)	<i>atpA</i> -back (0.35)
3	<i>glcDEF</i> (-0.07)	CO₂_ex (-0.21)	<i>cysC</i> (0.09)	H ₂ O outflow (0.22)
4	adhE (-0.05)	<i>mdh</i> (-0.17)	<i>mgsA</i> (0.09)	<i>nuoA</i> (0.14)
5	<i>ptaA</i> (-0.05)	H ₂ S_ex (-0.16)	<i>adk</i> (0.06)	<i>cysC</i> (0.11)
Lycopin				
	Knockout-Kandidaten		Überexpressionskandidaten	
	$\mathbb{S}_{\text{Lyc,R}}^{\text{aer}}$	$\mathbb{S}_{\text{Lyc,Y}}^{\text{aer}}$	$\mathbb{S}_{\text{Lyc,R}}^{\text{aer}}$	$\mathbb{S}_{\text{Lyc,Y}}^{\text{aer}}$
1	frd (-0.15)	<i>atpA</i> (-0.23)	DXP-Pathway (1.0)	DXP-Pathway (1.0)
2	<i>avtA</i> -back (-0.14)	<i>frd</i> (-0.21)	Lycopinsynthese (1.0)	Lycopinsynthese (1.0)
3	<i>nadB</i> (-0.14)	<i>tpi</i> -back (-0.18)	<i>avtA</i> (0.14)	tpi (0.18)
4	<i>glcDEF</i> (-0.10)	<i>sdh</i> (-0.16)	<i>pyrD</i> (0.09)	<i>avtA</i> (0.15)
5	<i>fdhF</i> (-0.09)	cyoA (-0.15)	<i>glcDEF</i> (0.08)	Succ_ex (0.15)

adk - Adenylat Kinase, *cysC* - Adenylsulfat Kinase, *hchA* - Glyox::Lactat, *glcDEF* - Glycolat Oxidase, *nadB* - Aspartat Oxidase, *fdhF* - Format Hydrogenlyase, *avtA* - Valine Pyruvat Aminotransferase, *pyrD* - Dihydroorotat Dehydrogenase

Die Übereinstimmung der *in silico* Interventionskandidaten mit experimentell bestätigten Interventionen zur Produktion von Lycopin ist für die Kandidaten besser, die mit dem GS-Modell identifiziert wurden (im Vergleich zu den mit dem ZSW-Modell identifizierten Kandidaten). Dies zeigt, dass die Verwendung genomskaliger Netzwerke von Vorteil für Produkte des Sekundärstoffwechsels sein kann. Insbesondere gilt dies für die Identifizierung von Interventionskandidaten, die nicht unmittelbar für die Synthese des Produktes von Bedeutung sind (und daher meist nicht Bestandteil der spezifischen ZSW-Modelle sind), aber dennoch Einfluss auf die Produktsynthese haben können. In Kombination mit CASOP wurde der EFMSampler in einer anderen Studie bereits erfolgreich zur Identifizierung von Knockouts zur Überproduktion von Aminosäuren mit *E. coli* genutzt (Pande et al., 2014). Der EFMSampler wurde hier zwar in Kombination mit CASOP demonstriert, seine möglichen Anwendungen sind aber nicht darauf beschränkt. Als naheliegende Anwendungen können z.B. die Ansätze der Minimalen Metabolischen Funktionalität oder der FluxDesign-Ansatz gewählt werden (siehe Abschnitt 4.4).

7 Statistische Betrachtung metabolischer Vergleichsstudien

In Kapitel 7.1 wird zunächst die Notwendigkeit statistischer Analysen für Vergleichsstudien motiviert. Die dafür benötigten mathematischen Grundlagen der statistischen Hypothesentests werden in Kapitel 7.2 erläutert. Anschließend werden in Kapitel 7.3 anhand einer Fallstudie die Auswirkungen verschiedener Varianzquellen metabolischer Raten auf die statistische Belastbarkeit von beobachteten Unterschieden systematisch analysiert. Das Vorgehen für eine statistische Versuchsplanung unter Verwendung etablierter Methoden der Statistik wird abschließend in Kapitel 7.4 diskutiert. Da diese meist nicht berücksichtigt werden, werden sie hier für bioprozesstechnische Anwendungen am Beispiel metabolischer Vergleichsstudien demonstriert.

7.1 Motivation

In der Prozessentwicklung einschließlich dem Metabolic Engineering unterliegen bestimmte Schlüsselgrößen einer ständigen Untersuchung und Optimierung. Eine Klasse solcher Schlüsselgrößen für Zellkultivierungen stellen die Stoffwechselflüsse der kultivierten Organismen dar. Um die für den jeweiligen Prozess relevanten Flüsse zu verbessern, werden die Auswirkungen veränderter Kultivierungsbedingungen auf die Stoffwechselflüsse in Vergleichsstudien analysiert. In diesen Studien werden zwei (oder mehrere) verschiedene Kultivierungsszenarien S_A und S_B definiert und anschließend die resultierenden Stoffflüsse bzw. Produktionsraten verglichen. Neue Kultivierungssysteme oder auch nur bestimmte Teilkomponenten eines Prozesses müssen einen signifikanten Vorteil für diese Produktionsraten gegenüber dem etablierten Prozess versprechen, um zeit- und kostenaufwändige Umstellungen zu rechtfertigen. Das gilt insbesondere für die Kultivierung tierischer Zelllinien, bei denen die experimentellen Kosten im Vergleich zu bakteriellen Kultivierungen sehr hoch sind. Für die meisten in der Literatur beschriebenen Vergleichsstudien wird pro experimentellem Setup jedoch nur ein Experiment oder bestenfalls bis zu drei Experimente pro Szenario durchgeführt. Ob mit diesen experimentellen Daten statistisch belastbare (d.h. signifikant im mathematischen Sinn) Aussagen getroffen werden können, wird bis auf wenige Ausnahmen dabei nicht untersucht (Carinhas et al., 2010). Eine statistische Analyse ist jedoch zwingend notwendig, um sowohl falsch-positive als auch falsch-negative Ergebnisse zu vermeiden. Daher wird in diesem Kapitel dargelegt, wie durch die Anwendung von grundlegenden statistischen Verfahren die Ergebnisse metabolischer Vergleichsstudien objektiv interpretiert werden können.

7.2 Mathematische Grundlagen

Mit Hilfe der statistischen Hypothesentests lassen sich beobachtete Unterschiede von metabolischen Raten unter zwei verschiedenen experimentellen Szenarien $\mathbb{S}_A$ und $\mathbb{S}_B$ objektiv analysieren. Die Definition geeigneter Effektgrößen (zum Beispiel der Unterschied der Raten bzw. deren Mittelwerte) sowie Annahmen über die Verteilung der Fehler oder Varianzen der Messgrößen bestimmen die Wahl des geeigneten Tests. Unabhängig vom gewählten Test muss vor Durchführung der statistischen Analysen ein Signifikanzniveau α festgelegt werden. Es gibt die obere Grenze der Irrtumswahrscheinlichkeit für einen Fehler 1. Art an, also der Wahrscheinlichkeit eine Beobachtung als signifikant zu beurteilen, obwohl die Nullhypothese korrekt ist.

7.2.1 Theorie statistischer Hypothesentests

Eine Klasse geeigneter statistischer Tests zur Analyse von Vergleichsstudien sind Tests auf Unterschiede im Lagemaß (wie zum Beispiel der Mittelwerte). In dieser Klasse von Tests sind wiederum verschiedene Testhypothesen denkbar. So lässt sich zum Beispiel ein zweiseitiges Testproblem formulieren als:

$$H_0 : \Theta = \Theta_0 \quad \text{vs.} \quad H_1 : \Theta \neq \Theta_0,$$

wobei Θ ein geeignetes Effektmaß, wie beispielsweise der Unterschied einer Rate in zwei Szenarien A und B darstellt ($\Theta = r_{i,A} - r_{i,B}$). Falls eine begründete Annahme eine gerichtete Hypothese zulässt, so ergeben sich für einen einseitigen Test die Hypothesen:

$$H_0 : \Theta \leq (\geq) \Theta_0 \quad \text{vs.} \quad H_1 : \Theta > (<) \Theta_0.$$

Im Folgenden wird das Testproblem einer einfachen ungerichteten Hypothese genauer untersucht. Es soll analysiert werden, ob sich die Mittelwerte metabolischer Flüsse zwischen zwei experimentellen Aufbauten signifikant unterscheiden. Mit $\Theta_0 = 0$ ist dies äquivalent zu:

$$H_0 : \overline{r_{i,A}} = \overline{r_{i,B}} \quad \text{vs.} \quad H_1 : \overline{r_{i,A}} \neq \overline{r_{i,B}}.$$

Da sich die Varianzen der Raten typischerweise aus der Summe normalverteilter Teilkomponenten zusammensetzen, ist es gerechtfertigt, die Normalverteilungsannahme auch für die Ratenwerte anzuwenden. Wird zusätzlich noch angenommen, dass die Gesamtvarianz der Raten in beiden Szenarien gleich groß ist, so kann ein t -Test für zwei unabhängige Stichproben verwendet werden. Je nach verwendetem Test sind verschiedene Voraussetzungen bzw. Annahmen an die Daten zu erfüllen. So kann mit als bekannt vorausgesetzten Varianzen auch ein Gauß-Test durchgeführt werden oder bei sehr sehr kleinen Stichproben ein Welch-Test, für den die Varianzen nicht bekannt sein müssen.

Da für diese Analysen die Ratenvarianz als bekannte Größe verwendet wird und weiterhin von normalverteilten Messwerten für die Raten ausgegangen wird, kann der Gauß-Test für zwei unabhängige Stichproben verwendet werden. Mit $\sigma_{i,A/B}^2$ als Varianz der Rate r_i und $n_{A/B}$ als Anzahl der Experimente in Szenario A bzw. B berechnet sich die zugehörige Teststatistik z durch:

$$z = \frac{\overline{r_{i,A}} - \overline{r_{i,B}} - \theta_0}{\sqrt{\frac{\sigma_{i,A}^2}{n_A} + \frac{\sigma_{i,B}^2}{n_B}}}. \quad (7.1)$$

Gleichung 7.1 kann erweitert werden, um auch die statistische Power $(1 - \beta)$ zu berücksichtigen. Wenn von derselben Anzahl Experimente für beide Szenarien ausgegangen wird ($n_A = n_B = n$), ergibt sich:

$$z_{1-\beta} = \frac{\sqrt{n} |\bar{r}_{i,A} - \bar{r}_{i,B} - \Theta_0|}{\sqrt{\sigma_A^2 + \sigma_B^2}} - z_{1-\alpha/2} \quad (7.2)$$

mit $z_{1-\alpha/2}$ und $z_{1-\beta}$ als den zugehörigen Quantilen der Standardnormalverteilung. Die angenommene Effektgröße ist durch $|\bar{r}_{i,A} - \bar{r}_{i,B} - \theta|$ gegeben.

Aus Gleichung 7.2 ist ersichtlich, dass die Teststatistik von den Varianzen der Raten abhängig ist. Da eine deutliche Erhöhung der Anzahl der Experimente aus Zeit- und Kostengründen meist nicht möglich ist und auch die Wahl des Signifikanzniveaus α sowie der Power $(1 - \beta)$ vor Beginn der Experimente festgelegt ist, spielen die Varianzen eine Schlüsselrolle für die Teststatistik. Auch um die Reproduzierbarkeit der Beobachtungen besser bewerten zu können, ist eine systematische Analyse der verschiedenen Quellen der Variabilität der berechneten Raten von besonderem Interesse.

7.2.2 Varianzquellen extrazellulärer Raten

Die Ursachen von Variationen bei der Bestimmung einer Austauschrate r_i können grundsätzlich in zwei verschiedene Klassen unterteilt werden. Zur ersten Klasse der Varianzquellen einer Rate, die als Methodenvarianz $\sigma_{M_i}^2$ bezeichnet wird, gehören alle Faktoren, die mit der beschränkten Messgenauigkeit der verwendeten Geräte oder der Messmethodik zusammenhängen. Als Beispiel seien hier die Messgenauigkeiten der Konzentrationsmessungen der ausgeschiedenen bzw. aufgenommenen Metabolite oder die Messgenauigkeit bei der Bestimmung der Zellzahl- bzw. Biomassemessungen genannt. Eine Möglichkeit, die Messvarianzen $\sigma_{C_i}^2$ der Konzentrationen C_i des über r_i ausgeschiedenen Metaboliten (oder ihre Standardabweichungen σ_{C_i}) abzuschätzen, besteht darin, jede Probe in mehrere Teilproben zu unterteilen und Mehrfachmessungen vorzunehmen (meist Dreifachmessungen). Dies ist jedoch nicht empfehlenswert, da selbst identische Wiederholungsmessungen (hohe Präzision) nicht zu der Interpretation führen dürfen, dass keine Messfehler vorliegen. Der wahre Datenwert kann auch in einem solchen Fall vom Messwert verschieden sein und somit ein Messfehler vorliegen. Ein aufwändigerer Zugang zur Bestimmung der Messgenauigkeiten ist die Validierung der Messmethoden bzw. Geräte (ICH Harmonised Tripartite Guideline 2005). Dies hat den Vorteil, dass die technischen Varianzen der Messmethoden $\sigma_{C_i}^2$ a priori berücksichtigt werden können und nicht über technische Mehrfachbestimmungen unzureichend abgeschätzt werden müssen. Der Einfluss dieser Varianzen auf die Genauigkeit der abgeleiteten Raten wurde bereits ausführlich untersucht (Bonarius et al., 1999; Goudar et al., 2009; Wiechert et al., 1997). Die Ungenauigkeiten in den Primärmessungen setzen sich je nach verwendetem Ansatz zur Ratenberechnung zu den Varianzen der Raten fort. Der explizite funktionale Zusammenhang $\sigma_{M_i}^2 = f(\sigma_{C_i}^2)$ dieser Fehlerfortpflanzung ist im Allgemeinen nicht zugänglich, kann aber mittels Monte-Carlo-Simulationen sehr genau abgeschätzt werden (Goudar et al., 2009). Zusammenfassend hängt die Methodenvarianz $\sigma_{M_i}^2$ einer Rate r_i somit ab von (i) der verwendeten Messtechnik (mit zugehöriger Messvarianz $\sigma_{C_i}^2$) und (ii) dem mathematischen Ansatz zur Berechnung der Rate r_i .

Die zweite Klasse von Varianzquellen, hier als Experimentalvarianz ($\sigma_{E_i}^2$) bezeichnet, wird nicht durch Messfehler verursacht. Sie quantifiziert verschiedene Quellen von Variationen zwischen verschiedenen Experimenten und kann spezifisch für einen Organismus und bzw. oder ein Kultivierungssystem sein. Ein Ansatz, diese Größe zu bestimmen, ist die Durchführung paralleler Experimente innerhalb eines Kultivierungssystems (technische Wiederholungen). Wenn dies der einzige Ansatz ist, so führt das jedoch zu einer Unterschätzung der biologischen Variabilität (die ein Hauptbestandteil der interexperimentellen Varianz ist) des verwendeten Organismus. Mögliche Ursachen biologischer Varianz wie z.B. copy-number-variations von Enzymen und anderen Molekülen (Elowitz et al., 2002; Rausenberger und Kollmann, 2008) oder stochastische Effekte werden in (Arriaga, 2009; Hilfinger und Paulsson, 2011) detailliert diskutiert. Selbst wenn Experimente unter exakt identischen Bedingungen wiederholt werden, können einige Größen (wie z.B. metabolische Raten) bei Verwendung unabhängiger Vorkulturen in einem erheblichen Ausmaß voneinander abweichen. Weitere Faktoren, die die Experimentalvarianz beeinflussen können, sind z.B. das Alter und der Zustand der Vorkultur, die Medium-Charge oder auch die durchführenden Personen. Die hier definierte Experimentalvarianz absorbiert auch weitere unspezifische Varianzquellen, wobei ein generell gültiger funktionaler Zusammenhang der verschiedenen Einflussgrößen auf die Experimentalvarianz nicht identifiziert werden kann.

Die Gesamtvarianz σ_i^2 einer einzelnen spezifischen Rate hängt von beiden beschriebenen Varianzkategorien ab:

$$\sigma_i^2 = g(\sigma_{M_i}^2, \sigma_{E_i}^2). \quad (7.3)$$

Der explizite funktionale Zusammenhang ist dabei im Allgemeinen nicht bekannt. Es kann jedoch angenommen werden, dass die methodische und biologische Varianz unabhängig voneinander sind. Weder beeinflusst die biologische Varianz die Genauigkeit der Konzentrationsmessungen noch haben Messvarianzen einen Einfluss auf die biologische Variabilität. Damit vereinfacht sich Gleichung 7.3 zu

$$\sigma_i^2 = \sigma_{M_i}^2 + \sigma_{E_i}^2. \quad (7.4)$$

Die Gesamtvarianz setzt sich also additiv aus den beiden diskutierten Teilkomponenten zusammen, wobei $\sigma_{E_i}^2$ gegebenenfalls weiter zerlegt werden könnte, um einzelne Effekte, die in dieser Darstellung kumuliert wurden, genauer zu betrachten.

7.3 Fallstudie zum Vergleich von Kultivierungssystemen

Als Fallstudie wird das Wachstumsverhalten der Vogel-Zelllinie AGE1.CR.pIX in zwei verschiedenen Kultivierungssystemen analysiert. Dabei soll untersucht werden, ob sich wichtige Stoffwechselraten der Zellen in Abhängigkeit vom Kultivierungssystem signifikant unterscheiden. Die Kultivierungen erfolgen entweder in einem STR- (stirred tank reactor) oder in einem Wave-Bioreaktor. Da sich die experimentellen Aufbauten sehr unterscheiden (z.B. bzgl. der Belüftung, Zellbewegung, pH-Kontrolle etc.), werden deutliche Unterschiede erwartet. Es gibt jedoch keine klare Hypothese, in welchem Kultivierungssystem die Zellen besser wachsen bzw. in welchem sie mehr Substrate aufnehmen oder Produkte bilden. Somit wird als zu widerlegende Nullhypothese angenommen, dass sich die Raten nicht zwischen den Kultivierungssystemen unterscheiden $H_0 : r_A = r_B$. Als Alternativhypothese wird angenommen, dass die entsprechenden Raten nicht identisch sind $H_1 : r_A \neq r_B$.

7.3.1 Experimentelle Daten

Zunächst werden die Austauschraten einiger Schlüsselmetabolite und deren zugehörige Gesamtvarianzen σ_i^2 bestimmt, die zur Beschreibung des Zellwachstums besonders geeignet sind. Die Konzentrationsprofile im betrachteten Zeitraum sind in Abbildung 7.1 dargestellt. Für die experimentellen Details der Konzentrationsmessungen sei auf (Hädicke et al., 2013) verwiesen.

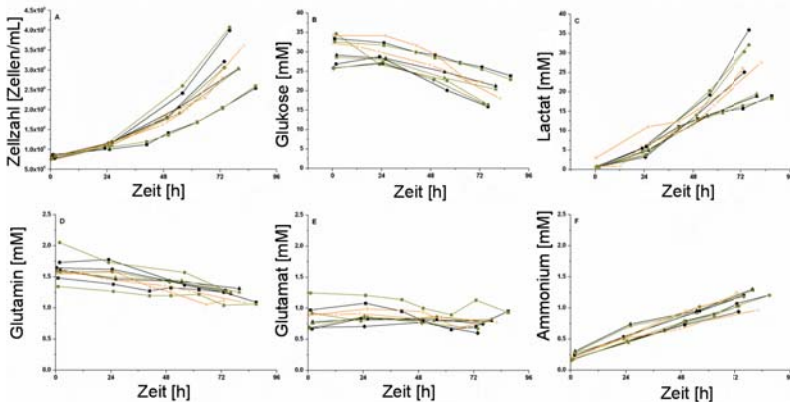


Abbildung 7.1: Konzentrationsprofile im beobachteten Zeitraum von der Zellzahl (a) sowie von b: Glukose, c: Lactat, d: Glutamin, e: Glutamat, f: Ammonium. Grüne und schwarze Linien: technische Wiederholungen (parallele Experimente) von biologisch unabhängigen STR-Experimenten. Orangene Linien: biologisch unabhängige Wave-Experimente.

Die Standardabweichungen der Konzentrationsmessungen sind durch die Verwendung validierter Messmethoden bekannt (Tabelle 7.1). Validierte Messmethoden haben den Vorteil (gegenüber Mehrfachmessungen einer Probe), dass die technische Varianz der Messmethoden berücksichtigt wird und die erwarteten Messgenauigkeiten *a priori* bekannt sind.

Tabelle 7.1: Standardabweichungen der Konzentrationsmessungen in absoluten Werten, wenn die Varianzen homogen im validierten Konzentrationsintervall sind; anderenfalls relativ bzgl des. Messwertes.

Metabolit	σ_C
Glukose	0.39 mM
Lactat	0.30 mM
Ammonium	4.5 %
Glutamin	12.8 %
Glutamat	0.03 mM
Zellzahl	2.5 %

Für jedes der biologisch unabhängigen Experimente werden die extrazellulären Raten und ihre zugehörigen empirischen Standardabweichungen aus den gemessenen extrazellulären Metabolitkonzentrationen berechnet (Tabelle 7.2). Für die STR-Experimente werden dabei die Mittelwerte aus zwei parallelen Kultivierungen verwendet. Die Varianzen der extrazellulären Raten hängen sowohl von der Genauigkeit der Konzentrationsmessungen als auch von den gemessenen Werten selbst ab. Daher ist es nicht möglich, für einen betrachteten Metaboliten *a priori* die zu erwartende Varianz der abgeleiteten Rate anzugeben. Dennoch können die empirischen Standardabweichungen der extrazellulären Raten mittels eines Monte-Carlo-Ansatzes bestimmt werden. Dafür werden für jedes Experiment 100.000 Datensätze generiert, wobei jeder Datenpunkt (Metabolit- und Zellkonzentrationen) mit einem normalverteilten Fehler $\epsilon N(0, \sigma_C)$ verrauscht wird. Für jeden simulierten Datensatz werden die Raten berechnet und anschließend die empirischen (relativen) Standardabweichungen σ_{M_i} berechnet. Die Mittelwerte $\bar{\sigma}_{M_i}$ der Standardabweichungen (über die Experimente) werden dann als Schätzer für die durchschnittlichen σ_{M_i} verwendet.

Um $\sigma_{E_i}^2$ zu schätzen, wird zunächst die Gesamtvarianz der Raten der vier unabhängigen STR-Experimente berechnet. Mit Hilfe der zuvor bestimmten durchschnittlichen Methodenvarianz σ_{M_i} und Gleichung 7.4 kann die interexperimentelle Varianz berechnet werden. Die in der Literatur oftmals berichtete Methodenvarianz $\sigma_{M_i}^2$ einer spezifischen Rate r_i kann meist nur als untere Schranke der Gesamtvarianz betrachtet werden. Sie erfasst jedoch nicht das gesamte Ausmaß der Varianz, die bei der Durchführung biologisch unabhängiger Experimente entsteht.

Die resultierenden Standardabweichungen der Raten sind in Tabelle 7.3 gegeben. Die großen relativen Gesamtstandardabweichungen für Glutamin und Glutamat sind nicht ungenauen Konzentrationsmessungen geschuldet (siehe Tabelle 7.1). Vielmehr sind die Konzentrationen annähernd konstant niedrig über den beobachteten Zeitraum. Daher sind die Ratenberechnungen sehr sensitiv bezüglich kleinsten Schwankungen. Im Extremfall absolut konstanter Konzentrationen ist die entsprechende Rate 0 und die zugehörige relative Varianz selbst für extrem genaue Messungen unendlich groß.

Tabelle 7.2: Berechnete Raten mit zugehörigen empirischen Standardabweichungen für STR- und Wave-Bioreaktor-Kultivierungen.

Fluss	STR 1	STR 2	STR 3	STR 4	Wave 1	Wave 2
Glukose	211.5 (14.5)	203.0 (17.9)	240.7 (16.7)	218.9 (15.3)	290.8 (15.3)	239.0 (15.6)
Lactat	308.5 (15.3)	387.2 (21.0)	604.6 (31.4)	524.4 (25.3)	390.3 (17.5)	433.9 (22.6)
Ammomium	18.7 (1.5)	15.9 (2.0)	13.9 (1.4)	8.7 (0.9)	15.6 (1.4)	9.2 (1.1)
Glutamin	5.9 (4.7)	5.5 (7.0)	10.7 (5.7)	6.0 (3.4)	8.7 (3.9)	7.6 (4.3)
Glutamat	11.0 (1.3)	0.8 (1.2)	1.2 (0.9)	2.1 (0.6)	6.2 (0.8)	2.5 (0.8)
μ	0.016 (6.0E-04)	0.018 (7.5E-04)	0.025 (8.9E-04)	0.021 (6.7E-04)	0.019 (6.1E-04)	0.022 (8.0E-04)

Wachstumsrate μ in h^{-1} , alle weiteren Raten in $\mu\text{mol}/(\text{gDW} \cdot \text{h})$.

Die interexperimentelle Varianz ist klar die Hauptquelle der Gesamtvarianz der berechneten Raten (außer für Glukose). Die Abweichungen zwischen den parallelen STR-Experimenten sind jeweils in der Größenordnung der Messgenauigkeit (nicht gezeigt). Wäre auf biologisch unabhängige Experimente verzichtet worden, hätte somit ausschließlich die Messungenauigkeit zu den Varianzen der Raten beigetragen. Dies entspricht jedoch nicht der Realität, da die Unterschiede zwischen biologisch unabhängigen Experimenten offensichtlich weitaus größer sind als nur durch die Messungenauigkeit erklärbar ist.

Tabelle 7.3: Empirische relative Standardabweichungen der Raten

Rate	Methoden-Std σ_M	Experiment-Std σ_E	Gesamt-Std σ
Glukose	6.9 %	3.4 %	7.7 % [4.2% ... 27.6%]
Lactat	5.0 %	28.8 %	29.2 % [16.5% ... 109%]
Ammomium	10.4 %	27.8 %	29.7 % [16.9% ... 111%]
Glutamin	74.3 %	-	34.2 % [19.4% ... 127%]
Glutamat	67.5 %	110.2 %	18.6 % [73.2% ... 482%]
μ	3.6 %	18.6 %	18.9 % [10.7% ... 70.6%]

Zur Gesamtstandardabweichung σ ist das zugehörige 95 % Konfidenzintervall angegeben.

Mit den Daten der vier STR-Experimente lassen sich die Gesamtvarianzen (bzw. Standardabweichungen) der Raten nicht sehr exakt abschätzen. Dies wird durch die großen 95% Konfidenzintervalle der zugehörigen Schätzer widerspiegelt (Tabelle 7.3). Die unteren Schranken dieser Konfidenzintervalle geben dennoch eine verlässliche Abschätzung der Mindestgröße der Varianzen. Im Folgenden werden diese Unterschranken als Annahme für die Gesamtvarianz der Raten verwendet. Damit ist das Risiko der Überschätzung der Gesamtvarianz sehr klein und diese Annahme eher als konservativ bzw. optimistisch zu bezeichnen.

7.3.2 Statistische Analyse der experimentellen Daten

Beim Vergleich einiger Flüsse genau eines STR-Experimentes mit genau einem zugehörigen Wave-Experiment ergeben sich zum Teil große Unterschiede. So unterscheiden sich z.B. die Raten von Glukose (für STR 2 vs. Wave 1), Lactat (für STR 3 vs. Wave 1) und Ammonium (für STR 4 vs. Wave 1) um 43%, 55% oder sogar 79%. Der Vergleich von Einzelexperimenten bestärkt somit die Hypothese signifikanter Unterschiede zwischen beiden Kultivierungssystemen. Jedoch fällt hier bereits auf, dass das Ausmaß der Differenzen sehr stark von den gewählten STR- und Wave-Experimenten abhängig ist.

Im nächsten Schritt werden nicht nur einzelne Experimente, sondern die Mittelwerte von jeweils zwei unabhängigen STR-Experimenten mit den Mittelwerten der zwei Wave-Experimente verglichen. Auch dabei ist es bis auf Glutamin für jeden Metaboliten möglich, eine Kombination von STR-Experimenten zu finden, so dass erneut deutliche Unterschiede beobachtet werden (Abbildung 7.2). Wie zuvor hängt das Ausmaß der Differenz deutlich von der gewählten Kombination der STR-Experimente ab.

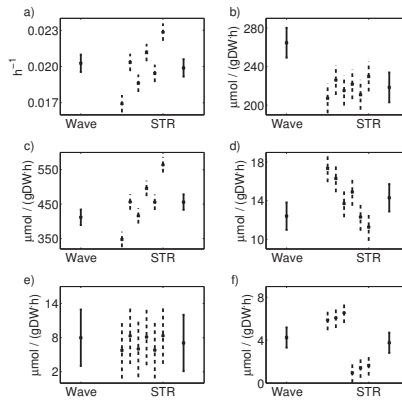


Abbildung 7.2: Mittelwerte der Austauschraten mit zugehöriger Standardabweichung (STD) der beiden Wave- und vier STR-Experimente (durchgezogene Linien). Gestrichelte Linien: Mittelwerte und STD von jeweils zwei STR-Experimenten (von links nach rechts: STR 1 und STR 2, STR 1 und STR 3, STR 1 und STR 4, STR 2 und STR 3, STR 2 und STR 4, STR 3 und STR 4). a) μ , b) Glukose, c) Lactat, d) Ammonium, e) Glutamin, f) Glutamat.

Da für jede Rate sechs voneinander abhängige Tests durchgeführt werden (sechs Möglichkeiten, zwei von vier verschiedenen STR-Experimenten auszuwählen), muss das Signifikanzniveau für multiples Testen korrigiert werden (hier wird die Bonferroni-Korrektur (Broadhurst und Kell, 2006) verwendet, was zu einem adjustierten Signifikanzniveau von $\alpha = 8.3E - 03$ führt). Um die Auswirkung der Varianzen bzw. Standardabweichungen auf das Testergebnis zu demonstrieren, werden im Folgenden zwei theoretische Szenarien analysiert. Im ersten Szenario $\mathbb{S}_1$ wird die interexperimentelle Varianz vernachlässigt, so dass die Gesamtvarianz gleich der Methodenvarianz ist (so wie es häufig in der Litera-

tur angewendet wird). Im zweiten Szenario $\mathbb{S}_2$ werden die zuvor bestimmten tatsächlichen unteren Schranken für die Gesamtvarianz verwendet. Da somit jeweils „bekannte“ Varianzen vorausgesetzt werden, kann ein z-Test gemäß Gleichung 7.1 verwendet werden.

Mit der Varianz von Szenario $\mathbb{S}_1$ resultiert mehr als ein Drittel aller Tests (14 von 36) in signifikanten Unterschieden (siehe Anhang A.2). Wird hingegen die größere Gesamtvarianz von Szenario $\mathbb{S}_2$ verwendet, so reduziert sich die Anzahl signifikanter Tests auf drei. Es stellt sich somit die Frage, ob die größere Varianz falsch-positive Ergebnisse eliminiert oder falsch-negative Ergebnisse zur Folge hat.

Um diese Frage zu beantworten, werden zusätzlich die drei möglichen Zweierkombinationen (1,2 vs. 3,4 ; 1,3 vs. 2,4 und 1,4 vs. 2,3) von STR-Experimenten getestet. Hier wird das Signifikanzniveau mittels Bonferroni-Korrektur auf $\alpha = 1.7E - 02$ angepasst. Es zeigt sich, dass mit der Varianz von Szenario $\mathbb{S}_1$ die Hälfte aller Tests (neun von 18) einen signifikanten Unterschied detektieren (siehe Anhang A.2). Dieses sind jedoch falsch-positive Ergebnisse, da die Raten alle zur selben Grundgesamtheit (dem STR-Kultivierungssystem) gehören. Dass auch mit der Varianz aus Szenario $\mathbb{S}_2$ noch drei signifikante Ergebnisse erzielt werden, kann dadurch erklärt werden, dass lediglich untere Schranken für die Gesamtvarianzen angenommen werden und somit die tatsächliche Varianz unterschätzt wird.

Im letzten Schritt werden alle verfügbaren Daten verwendet und die Mittelwerte der Flüsse aller STR-Experimente mit den Mittelwerten der Flüsse der Wave-Experimente verglichen. Die Raten von Lactat, Ammonium und Glukose unterscheiden sich zwar nicht mehr so drastisch wie zuvor, dennoch betragen die Unterschiede 11%, 15% bzw. 21%. Für die statistischen Tests wird die Varianz nun nicht mehr als bekannt vorausgesetzt, sondern aus den Daten geschätzt. Damit kann der klassische t-Test verwendet werden, um die Signifikanz der beobachteten Unterschiede zu bewerten.

Tabelle 7.4: p -Werte und 95% Konfidenzintervalle der Differenz der Mittelwerte der Raten unter Verwendung aller Daten.

Rate	p -Werte	Konfidenzintervall
Glukose	0.08	$[-9.1 \dots 101.8]$
Lactat	0.68	$[-235.9 \dots 324.0]$
Ammonium	0.64	$[-8.5 \dots 12.3]$
Glutamin	0.64	$[-4.1 \dots 6.0]$
Glutamat	0.90	$[-10.0 \dots 11.0]$
μ	0.91	$[-8.7E - 03 \dots 8.0E - 03]$

Die resultierenden p -Werte der Tests liegen zwischen 0.08 und 0.91, so dass für keine Rate der Unterschied zwischen den Kultivierungssystemen als signifikant bezeichnet werden kann. Entsprechend beinhalten die zugehörigen 95%-Konfidenzintervalle der Differenzen der Ratenmittelwerte die Null. Entgegen der ursprünglichen Annahme kann mit den gegebenen Daten die Nullhypothese nicht verworfen werden (Tabelle 7.4). Diese Ergebnisse verdeutlichen die Notwendigkeit biologisch unabhängiger Wiederholungsexperimente, um die reale Gesamtvarianz korrekt zu bestimmen. Wird die Gesamtvarianz unterschätzt, so steigt die Rate von falsch-positiven Ergebnissen schnell an. Nur wenn die verwendete Varianz der realen Varianz entspricht, kann das korrekte Signifikanzniveau eingehalten werden (Abbildung 7.3).

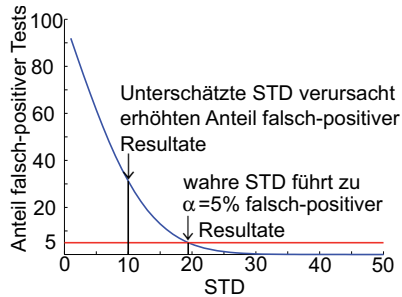


Abbildung 7.3: Anteil falsch-positiver Tests in Abhängigkeit der Standardabweichung (STD).

7.4 Statistische Versuchsplanung

Eine wichtige Fragestellung der statistischen Versuchsplanung ist es, zu bestimmen, ob die erwarteten experimentellen Vorteile mit ausreichender statistischer Evidenz belegt werden können. Eine Analyse der statistischen Power der experimentellen Studie (Fitzner and Heckinger 2010) ermöglicht es, die Abhängigkeiten zwischen

1. der (standardisierten) Effektgröße,
2. der Fallzahl $N = n_A + n_B$,
3. des Signifikanzniveaus α und
4. der Power des Tests ($\pi = 1 - \beta$) (die Wahrscheinlichkeit H_0 abzulehnen, wenn H_1 wahr ist)

zu untersuchen. In der konkreten Anwendung entsprechen die ersten beiden Punkte dem zu zeigenden Unterschied der Raten sowie der Anzahl der biologisch unabhängigen Experimente. Um eine Fallzahlplanung durchführen zu können, müssen vor der Durchführung der Experimente sowohl

- das Signifikanzniveau α (um die Falsch-Positiv-Rate zu kontrollieren) als auch
- die gewünschte Power der Studie (die Wahrscheinlichkeit, einen vorhandenen Unterschied auch zu detektieren)

festgelegt werden. Zwei weitere Größen, deren genaue Werte *a priori* nur selten zur Verfügung stehen, die aber für die Versuchsplanung von Bedeutung sind, sind

- die erwarteten und bzw. oder vorhandenen Ratenunterschiede sowie
- die Varianzen der berechneten Raten.

Gleichung 7.2 kann dann genutzt werden, um die benötigte Anzahl Experimente zu berechnen. Alternativ können (bei gegebener Fallzahl und Effektgröße) obere Schranken für die Varianzen berechnet werden, so dass die Beobachtung zu einem positiven Testergebnis führt (H_0 wird verworfen).

Um Aussagen unabhängig von konkreten Messwerten treffen zu können, werden im Folgenden sowohl der Mittelwert von $r_{i,B}$ als auch die Varianzen (bzw. Standardabweichungen) relativ zum Mittelwert von $r_{i,A}$ betrachtet:

$$\begin{aligned}(\bar{r}_{i,B}) &= (1 + a) \cdot (\bar{r}_{i,A}), \\ \sigma_{i,A} &= b \cdot (\bar{r}_{i,A}), \\ \sigma_{i,B} &= (1 + a) \cdot b \cdot (\bar{r}_{i,A}),\end{aligned}\tag{7.5}$$

wobei a als angenommener real vorhandener relativer Unterschied der Mittelwerte interpretiert werden kann und b die relative Größe der Standardabweichung beschreibt. Mit diesen Standardisierungen kann Gleichung 7.2 umformuliert werden, um die akzeptable Größe der relativen Standardabweichung zu berechnen:

$$b = \frac{(\sqrt{n} |a|)}{\left((z_{(1-\alpha/2)} + z_{(1-\beta)}) \sqrt{(1 + (1 + a)^2)}\right)}.\tag{7.6}$$

Alternativ ist auch eine Umformulierung möglich, um die benötigte Fallzahl zu berechnen:

$$n = \left(\frac{(b(z_{(1-\alpha/2)} + z_{(1-\beta)}) \sqrt{(1 + (1 + a)^2)})^2}{(|a|)} \right).\tag{7.7}$$

In den Gleichungen 7.6 und 7.7 wechselt die Zuordnung der abhängigen und unabhängigen Variablen. So werden in Gleichung 7.6 für gegebene Fallzahlen die zugehörigen Standardabweichungen berechnet. In Gleichung 7.7 hingegen kann für gegebene Standardabweichungen die Fallzahl berechnet werden. Nachfolgend werden die Gleichungen 7.6 und 7.7 genutzt, um die Zusammenhänge zwischen a , b und n für gegebenes $\alpha = 0.05$ und $\beta = 0.2$ zu analysieren.

Die Fallstudie des vorigen Abschnitts hat gezeigt, dass selbst bei deutlichen Unterschieden der Raten keine statistische Signifikanz erreicht wird, wenn die Gesamtvarianz der Raten entsprechend groß ist. Somit sollte stets vor Durchführung der Experimente überprüft werden, ob der erwartete Effekt mit der zur Verfügung stehenden Anzahl an Experimenten belegt werden kann. Ebenso sollte vor der Durchführung der Experimente geklärt werden, ob eine ein- oder zweiseitige Hypothese aufgestellt wird, da auch dies einen erheblichen Einfluss auf die benötigte Fallzahl hat.

Die Abhängigkeit der Anzahl benötigter biologisch unabhängiger Experimente vom relativen Unterschied der durchschnittlichen Raten und der relativen Gesamtstandardabweichung ist in Abbildung 7.4 dargestellt. Wenn die Standardabweichung 20% übersteigt oder der Unterschied der Raten kleiner als 50% ist, sind die üblichen zwei unabhängigen Experimente pro Szenario nicht mehr ausreichend.

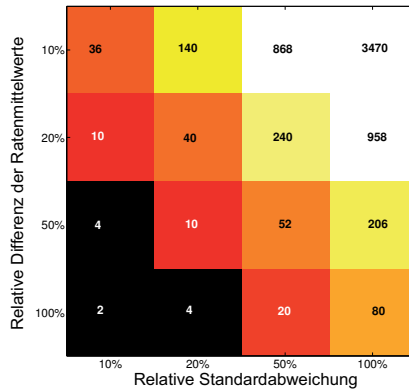


Abbildung 7.4: Anzahl benötigter Experimente in Abhängigkeit der durchschnittlichen Raten und der relativen Gesamtstandardabweichung.

Weitere Experimente sind notwendig, wenn an Stelle eines grundsätzlichen Unterschiedes ein gerichteter Mindestunterschied in den Raten (wie z.B. $r_{i,A} - r_{i,B} > 0.25r_{i,A}$) belegt werden soll. Es sei beispielsweise angenommen, dass ein realer Effekt zwischen zwei Bedingungen existiert, so dass ein Unterschied von 50 % in den Raten beobachtet werden kann (z.B. $r_A = 100, r_B = 150$). Zusätzlich sei die relative Gesamtstandardabweichung der Raten mit 25 % gegeben. Um mit statistischer Sicherheit schlussfolgern zu können, dass $r_B - r_A \geq 25$ (mit $\alpha = 0.05$ und $\pi = 0.8$) gilt, müssten mindestens sieben biologisch unabhängige Experimente pro Szenario durchgeführt werden.

Extrazelluläre Raten dienen als Eingangsvariablen für die Stoffflussanalyse, um intrazelluläre Flussverteilungen zu berechnen. Die Fortpflanzung von Messfehlern der extrazellulären Raten zu den abgeleiteten intrazellulären Flüssen (z.B. via MFA) wurde z.B. in (Carinhas et al., 2010, 2011; Goudar et al., 2009; Llaneras und Picó, 2007; van der Heijden et al., 1994) untersucht. Wenn die Varianzen der intrazellulären Raten bestimmt werden können, ist die hier präsentierte statistische Analyse analog für Vergleiche von intrazellulären Flüssen anwendbar.

Das hier verwendete Signifikanzniveau $\alpha = 0.05$ und die gewünschte Power $\beta = 0.8$ sind keine axiomatischen Werte. In konkreten Anwendungen können auch weniger stringente Signifikanzniveaus (wie z.B. $\alpha = 0.1$) genügen oder auch eine höhere statistische Power erforderlich sein.

Zusammenfassend zeigt die vorangegangene Fallstudie, dass die interexperimentelle Varianz eine Hauptquelle der Gesamtvarianz einer metabolischen Rate darstellen kann. Es ist mithin zwingend erforderlich, dass sie bei der Bestimmung der Gesamtvarianz berücksichtigt wird, um falsch-positive Ergebnisse zu vermeiden. In Bezug auf die Kosten sowie den zeitlichen und experimentellen Aufwand von Zellkultivierungen ist es somit sehr empfehlenswert, das Studiendesign *a priori* im Hinblick auf die statistische Aussagekraft sorgfältig zu analysieren.

8 Experimentelle Überprüfung einer ausgewählten Intervention

In diesem Kapitel wird exemplarisch eine im vorherigen Kapitel mittels CASOP identifizierte Intervention experimentell überprüft (von anderen Arbeitsgruppen wurde CASOP bereits erfolgreich eingesetzt (Yang et al., 2013; Pande et al., 2014)). In der Fallstudie zur Produktion von Succinat und Lactat mit *E. coli* wurde durch CASOP eine gesteigerte ATP-Hydrolyse als geeigneter Interventionskandidat identifiziert. Für beide Produkte wurden eine ATP-Quelle als Knockout- sowie eine ATP-Senke als Überexpressionskandidat identifiziert (Tabellen 6.4 und 6.6). Daher soll im Folgenden der Einfluss einer gesteigerten ATP-Hydrolyse auf die Produktion von Succinat und Lactat experimentell untersucht werden. Da diese Intervention unter aeroben Bedingungen bereits untersucht wurde, wird der Fokus hier auf die Auswirkungen für anaerobe Bedingungen gelegt.

8.1 Manipulation von Co-Faktoren

Die Co-Faktor-Regenerierung ist bei der Entwicklung mikrobieller Produktionsprozesse von entscheidender Bedeutung (Weusthuis et al., 2011). Besonders die Redoxbilanz sowie die Frage nach der Art der Prozessführung (aerob oder anaerob) ist für die realisierbaren Produktausbeuten maßgebend. Die Manipulation von Co-Faktoren als Werkzeug des Metabolic Engineering mit *E. coli* wurde bereits in früheren Studien experimentell untersucht (San et al., 2002; Holm et al., 2010). Das Hauptaugenmerk dieser Studien liegt dabei auf dem Einfluss verschiedener Verfügbarkeiten von NAD(P)H und ATP. So konnte eine weitere Studie zeigen, dass eine Verbesserung der NADPH-Verfügbarkeit (durch Erhöhung der intrazellulären NADPH-Konzentration) einen positiven Einfluss auf die Produktion von Polyphenolen hat (Chemler et al., 2010).

Die Auswirkungen eines geänderten ATP-Niveaus werden kontrovers diskutiert. Verschiedene Studien berichten gegensätzliche Beobachtungen bezüglich des Nutzen oder Schaden erhöhter ATP-Niveaus für die Produktivität bestimmter Substanzen. Einerseits haben verschiedene Studien gezeigt, dass hohe ATP-Niveaus zu schnellen Wachstumsraten und erhöhten Produktivitäten von Succinat und rekombinanten Proteinen führen können (Singh et al., 2011; Kim et al., 2012a). Andererseits bewirkt eine Reduzierung des ATP-Niveaus mittels eines ATP verbrauchenden Zyklus (engl. futile cycle) (Patnaik et al., 1992; Chao et al., 1993) oder ATP-Hydrolyse (Koeblmann et al., 2002) ebenfalls eine Erhöhung der Ausscheideraten von fermentativen Produkten sowie der Glukoseaufnahmerate. In Übereinstimmung zu den erstgenannten Studien wurde bei niedrigeren ATP-Niveaus eine reduzierte Wachstumsrate beobachtet. Die genannten Studien erfolgten

dabei jedoch ausschließlich unter aeroben Bedingungen.

8.2 ATP-Verbrauch zur Steigerung der Produktivität

Die Beziehung zwischen optimaler Ausbeute und optimaler Produktivität wurde in Bezug zum Zellwachstum bereits detailliert analysiert (von Stockar, 2013). Es kann gezeigt werden, dass für ausbeuteoptimale Flusslagen im Extremfall keine Antriebskraft für den Pathway mehr vorhanden sein kann. Dieser würde somit unendlich langsam ablaufen. Im anderen Extrem wird zwar ein sehr schneller Fluss realisiert, für den jedoch die Ausbeute gegen Null tendiert.

Eine analoge Argumentation wird benutzt, um die optimale (im Laufe der Evolution etablierte) ATP-Menge zu bestimmen, die entlang eines Pathways gewonnen werden kann (Werner et al., 2010). Dabei zeigt sich, dass nur ein Teil der verfügbaren freien Energie (ΔG) für die ATP-Ausbeute benutzt wird. Der verbleibende Teil der freien Energie wird benötigt, um den Pathway mit einer gewissen Rate zu betreiben. Um die optimale ATP-Produktivität entlang eines Pfades zu gewährleisten, muss jeweils die Hälfte der freien Energie des Substrates in die ATP-Synthese und in die Antriebskraft für den Pfad investiert werden. Diese Argumentation lässt sich für die Rate beliebiger Pathways erweitern. Um einen ausreichenden Fluss zu gewährleisten, muss ein Teil der verfügbaren freien Energie für die Generierung der Antriebskraft verwendet werden. Dies erfolgt zu Lasten der ATP-Ausbeute, da mit jedem hydrolysierten (bzw. nicht gewonnenen) ATP entlang eines Pfades dessen $\Delta_p G$ um etwa -35 bis -50 kJ/mol verringert wird. Wird entlang eines Pfades durch eine artifizielle Intervention (z.B. mittels futile cycling) zusätzliches ATP hydrolysiert, so wird eine höhere Antriebskraft generiert und damit dessen potentieller Umsatz erhöht.

Kopplung von ATP-Synthese und Zellwachstum

Die vorangegangene Argumentation wird hier erweitert, um einen erhöhten ATP-Bedarf als allgemeines Prinzip zur Steigerung der Ausbeute und Produktivität zu begründen. Durch die geringe Effizienz der ATP-Generierung mittels fermentativer Pathways werden nur geringe Mengen ATP für das Zellwachstum bzw. den Erhaltungsstoffwechsel bereit gestellt. Es muss somit ein hoher Fluss von Substrat zu fermentativem Produkt etabliert werden, um die benötigten Mengen ATP generieren zu können (Abbildung 8.1). Dies verursacht unter anaeroben Bedingungen generell höhere erreichbare Produktausbeuten (Weusthuis et al., 2011). Wird nun zusätzlich ATP abgezogen, müssen die Flüsse wie folgt reorganisiert werden:

Im ungünstigsten Fall wird die Substrataufnahmerate reduziert, was eine geringere Wachstumsrate und Produktivität zur Folge hat. Dennoch muss das Verhältnis der Flüsse zu Gunsten der Pfade verschoben werden, entlang derer die ATP-Synthese stattfindet, was zu einer gesteigerten Produktausbeute führt. Wenn die Aufnahmerate konstant bleibt, verschiebt sich wie zuvor das Verhältnis der Flüsse zu Gunsten der Pfade, entlang derer eine ATP-Synthese stattfindet. In diesem Szenario ist zwar die Wachstumsrate reduziert, jedoch ist sowohl eine gesteigerte Produktivität als auch eine verbesserte Produktausbeute die Folge. Im günstigsten Fall wird die Substrataufnahmerate erhöht, um die abgezogene

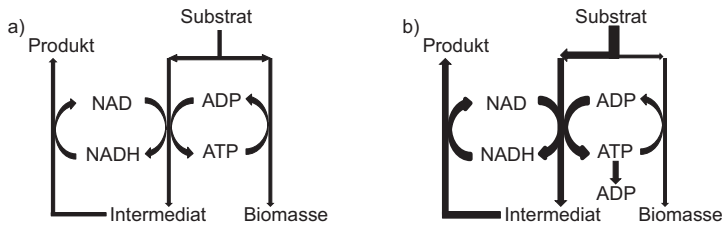


Abbildung 8.1: Kopplung von ATP-Synthese und Biomassewachstum. a) Wildtyp und b) mit künstlicher ATP-Senke. Im günstigsten Fall wird die Aufnahme rate erhöht, um die abgezogene ATP-Menge zu ersetzen und so zusätzliches Produkt gebildet.

ATP-Menge durch einen verstärkten Fluss durch die ATP-synthetisierenden Pathways zu ersetzen. In diesem Fall bleibt zum einen die Wachstumsrate konstant und des Weiteren erhöhen sich Ausbeute und Produktivität (Abbildung 8.1).

8.3 Experimentelle Durchführung

8.3.1 Überexpression von *ppsA*

Wie oben bereits erwähnt, wurde in den in der Literatur verfügbaren Studien der ATP-Haushalt ausschließlich unter aeroben Bedingungen manipuliert. Dafür wird entweder die membrangebundene F1-ATPase überexprimiert (Koebsmann et al., 2002) oder es werden verschiedene futile cycles implementiert (Patnaik et al., 1992; Chao et al., 1993). Eine erhöhte ATP-Hydrolyse unter anaeroben Bedingungen zu implementieren, ist eine experimentelle Herausforderung. Bisher ist in der Literatur keine Studie verfügbar, in der diese Manipulation erfolgreich beschrieben wurde. Daher wird im Folgenden ein erhöhter ATP-Verbrauch über eine Überexpression des endogenen *ppsA*-Gens realisiert. Dieses Gen kodiert für das Enzym Phosphoenolpyruvat Synthetase, welche Pyruvat und ATP zu PEP, AMP und Pi umsetzt. AMP wird dann über die ATP-AMP Transphosphorylase mit einem weiteren ATP wieder zu ADP phosphoryliert. Dieser Schritt wurde im Modell zusammengefasst zu:



Unter der Annahme, dass ein erhöhtes Level dieses Enzyms zu einem erhöhten Fluss von Pyruvat zu Phosphoenolpyruvat entgegen der Flussrichtung der Glykolyse führt, wird dadurch ein futile cycle implementiert, der netto ATP verbraucht und somit als ATP-Senke fungiert. Eine frühere Studie zeigt, dass die Überexpression von *ppsA* unter aeroben Bedingungen zu einer Stimulierung des Glukose-Stoffwechsels führt (Patnaik et al., 1992). Dies ist ein weiteres Indiz dafür, dass *ppsA* als Überexpressionskandidat auch unter anaeroben Bedingungen ein geeigneter Interventionskandidat ist. Zudem wurden sowohl die ATPdrain Reaktion als auch die Phosphoenolpyruvat Synthetase (*ppsA*) durch CASOP

für die Produktion von Succinat und Lactat als Überexpressionskandidaten identifiziert. Daher wird im Folgenden der Einfluss der Überexpression von *ppsA* auf die Produktionsraten für diese beiden Produkte untersucht.

8.3.2 Kultivierungen

Als Produktionsstamm wurde der *E. coli* Stamm KBM10111 gewählt. In diesem ist die Ethanol- und Acetatausscheidung durch Deletion der Alkoholdehydrogenase (*adhE*) und Phosphotransacetylase (*pta*) unterdrückt. Die Überexpression von *ppsA* wurde durch die Integration des Plasmids *pRR48-ppsA* gewährleistet. Auf diesem ist neben einer Resistenz für das Antibiotikum Ampicillin ein durch IPTG induzierbarer Promotor sowie das *ppsA*-Gen kloniert.

Die Kultivierungen erfolgten in Kolben in TANAKA-Medium mit einer Glukosestartkonzentration von 4 g/l und einer Biomassestartkonzentration von 0.044 g/l ($OD_{600} = 0.2$). Die Induktion der Zellen erfolgte durch Zugabe von IPTG mit einer Konzentration von 100 $\mu\text{mol/l}$ zum Zeitpunkt t_0 . Die Konzentrationen von Succinat, Lactat und Glukose wurden mittels Enzymtests bestimmt. Zur Kontrolle wurden zudem Format, Acetat und Ethanol gemessen (mittels Enzymtests). Die Biomassekonzentration wurde mittels optischer Dichtebestimmung bei 420 nm gemessen. Aus Vorversuchen ist die Messgenauigkeit der Enzymtests sowie der Biomassebestimmung bekannt ($\leq 5\%$). Um den Einfluss der biologischen Varianz zu minimieren, wurden die Zellen für jeweils eine Kultivierung mit und ohne IPTG aus denselben Vorkulturen gewonnen. Diese beiden Kultivierungen wurden parallel in einer technischen Doppelbestimmung durchgeführt. Dieses Experiment wurde zudem einmal biologisch unabhängig wiederholt, so dass insgesamt acht Kultivierungen durchgeführt werden.

8.3.3 Experimentelle Ergebnisse

Die Zellen zeigten während der ersten acht Stunden das erwartete exponentielle Wachstumsverhalten. Ab diesem Zeitpunkt stoppte das Zellwachstum jedoch sowohl für die induzierten als auch für die uninduzierten Zellen (Daten nicht gezeigt). Da für diese späten Zeitpunkte die Annahme des Fließgleichgewichtes nicht mehr erfüllt ist, wird die Analyse der experimentellen Daten auf die ersten acht Stunden begrenzt.

Die Induktion zeigt keinen Einfluss auf das Zellwachstum, da für induzierte und uninduzierte Zellen annähernd identische Biomassekonzentrationen beobachtet werden können (Abbildung 8.2). Induzierte Zellen nehmen innerhalb der ersten acht Stunden Glukose deutlich schneller als uninduzierte Zellen auf. Entsprechend führt die Induktion von *ppsA* zu der gewünschten erhöhten Ausscheidung von Succinat und Lactat. Die resultierende Lactatausbeute von 1.64 (mol Lactat / mol Glukose) mit Induktion ist im Vergleich zur molaren Ausbeute von 1.55 (mol Lactat / mol Glukose) ohne Induktion um 6% erhöht. Für Succinat ergibt sich durch die Induktion ebenfalls eine geringe Steigerung der Ausbeute von 0.09 mol auf 0.10 mol pro mol Glukose.

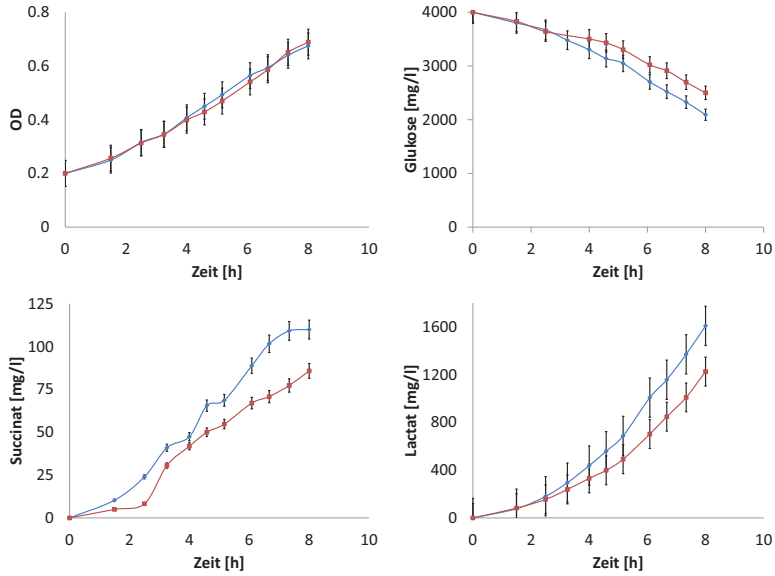


Abbildung 8.2: Mittelwerte (über alle Experimente) und zugehörige Standardabweichungen der Konzentrationen von Biomasse, Glukose, Succinat und Lactat. Rot: uninduzierte Zellen, blau: induzierte Zellen.

8.4 Analyse der experimentellen Ergebnisse

8.4.1 Bestimmung der Raten und Flussverteilungen

Die Wachstumsrate μ wird mittels der Methode der kleinsten Quadrate bestimmt. Dafür werden die gemessenen Biomassekonzentrationen $X(t)$ an eine exponentielle Wachstumsfunktion angepasst:

$$X(t) = X(t_0) \cdot e^{\mu \cdot t}. \quad (8.1)$$

Ebenso werden die Raten r_i für Glukoseaufnahme bzw. Lactat- und Succinatausscheidung (in [mmol/(gDW·h)]) durch die Methode der kleinsten Quadrate geschätzt. Die Anpassung der Konzentrationsdaten erfolgt dabei an folgende Funktion (mit μ aus Gleichung 8.1):

$$C_i(t) = C_i(t_0) + r_i \cdot \frac{X(t_0)}{\mu} \cdot (e^{\mu \cdot t} - 1). \quad (8.2)$$

Die Unterschiede der bestimmten Raten der Kultivierungen mit bzw. ohne Induktion von *ppsA* werden anschließend auf ihre statistische Signifikanz untersucht. Dazu wird zunächst die Methodenvarianz $\sigma_{M_i}^2$ einer Rate r_i durch Monte-Carlo-Simulationen geschätzt (wie in Kapitel 7.3.1 beschrieben). Da die induzierten und uninduzierten Zellen jeweils aus identischen Vorkulturen stammen, ist der Anteil der biologischen Varianz auf

Grund unabhängiger Vorkulturen an der Experimentalvarianz hier nicht vorhanden. Dennoch wird für die Experimentalvarianz $\sigma_{E_i}^2$ eine relative Standardabweichung von 10% angenommen, da sie weitere unspezifische Varianzquellen integriert. Die Gesamtvarianz der Raten ist damit gegeben durch: $\sigma_i^2 = \sigma_{M_i}^2 + \sigma_{E_i}^2$. Die Gesamtvarianz der Raten wird mithin als bekannt angenommen, so dass ein Zweistichproben-Gauß-Test für abhängige (verbundene) Stichproben angewendet werden kann. Es wird jeweils eine einseitige Hypothese (für alle Raten r_i) mit $H_0 : r_{i,\text{induziert}} > r_{i,\text{uninduziert}}$ gegen $H_1 : r_{i,\text{induziert}} \leq r_{i,\text{uninduziert}}$ zum Signifikanzniveau $\alpha = 0.05$ getestet.

Abschließend werden die intrazellulären Flussverteilungen der induzierten gegenüber den uninduzierten Zellen verglichen. Da die gemessenen Raten nicht ausreichen, um die Flussverteilungen mittels MFA eindeutig zu berechnen (siehe Kapitel 4.2.2), werden weitere Nebenbedingungen getroffen. Nach der Inaktivierung der Reaktionen *glkA* (Glukoseaufnahme via *ptsA*), *sdh* (nur aerob aktiv), *mae* (nur aerob aktiv), *poxB* (keine Acetatausscheidung beobachtet), *iclR* (nur mit Acetat als Substrat aktiv), *zwf*, *edd*, *ppc*, *mgsA*, *ppsA* und der CO₂-Ausscheidung kann die Flussverteilung der uninduzierten Zellen bestimmt werden. Die hier relevante Veränderung des ATPdrain-Flusses wird durch diese Annahmen nicht beeinflusst (siehe unten). Der resultierende ATPdrain-Fluss wird anschließend als Nebenbedingung zur Berechnung der Flussverteilung mit Induktion verwendet.

8.4.2 Auswertung der Raten und Flussverteilungen

Die durchschnittlichen resultierenden spezifischen Aufnahme- und Produktionsraten sind in Tabelle 8.1 gegeben. Unter induzierten Bedingungen ist die Glukoseaufnahmerate um 35.5 % gegenüber uninduzierten Bedingungen erhöht. Die Lactat- und Succinat-Produktionsraten sind ebenfalls um 35.8 % bzw. 37.3 % erhöht. Die Unterschiede der Austauschraten werden auf dem Signifikanzniveau von $\alpha = 5\%$ als signifikant eingestuft. Diese Ergebnisse sind sensitiv bezüglich der angenommenen relativen biologischen Varianz. Wird σ_{E_i} von 10 % auf 20 % verdoppelt, so ergeben sich für die Austauschraten von Glukose, Succinat und Lactat entsprechende p -Werte von 0.088, 0.074 und 0.070. Damit würden die beobachteten Unterschiede knapp nicht mehr als statistisch signifikant klassifiziert (mit dem zuvor definierten α). Um zu beurteilen, ob die verwendete Experimentalvarianz repräsentativ ist, sind zusätzliche Experimente zur Bestimmung der spezifischen Experimentalvarianz des hier verwendeten *E. coli*-Stammes notwendig, die im Rahmen dieser Arbeit nicht mehr durchgeführt werden können.

Tabelle 8.1: Wachstumsrate μ (in h⁻¹) und Austauschraten für Glukose, Succinat und Lactat (in mmol/(gDW h)) sowie zugehörige Standardabweichungen σ_{M_i} , σ_{E_i} und σ_i . p -Wert des entsprechenden einseitigen Gauß-Tests.

	Uninduziert		Induziert		p
	Rate	$\sigma_{M_i}/\sigma_{E_i}/\sigma_i$	Rate	$\sigma_{M_i}/\sigma_{E_i}/\sigma_i$	
μ	0.161	0.003/0.016/0.016	0.160	0.003/0.016/0.016	0.525
Glukose	-11.0	1.2/1.1/1.63	-14.9	1.2/1.49/1.91	0.014
Succinat	1.02	0.1/0.1/0.14	1.4	0.1/0.14/0.17	0.008
Lactat	19.0	0.7/1.9/2.02	25.8	0.9/2.58/2.73	0.002

Die resultierenden intrazellulären Flusslagen sind qualitativ sehr ähnlich. Sie unterscheiden sich hauptsächlich durch eine größere Glukoseaufnahme verbunden mit einem erhöhten glykolytischen Fluss sowie Lactatausscheidung (Abbildung 8.3). Die Induktion von *ppsA* bewirkt somit einen Fluss der Phosphoenolpyruvat Synthetase und einen damit erhöhten ATP-Verbrauch von 6.9 mmol/(gDW h). Dieser Wert für den erhöhten ATP-Verbrauch kann auch ohne die zusätzlich getroffenen Annahmen der MFA bestimmt werden. Werden ausschließlich die gemessenen Raten als Nebenbedingungen benutzt und eine FVA-Analyse durchgeführt, so ergeben sich für die oberen Schranken des ATPdrain-Flusses 13 bzw. 20 mmol/(gDW h) entsprechend ohne und mit Induktion von *ppsA*.

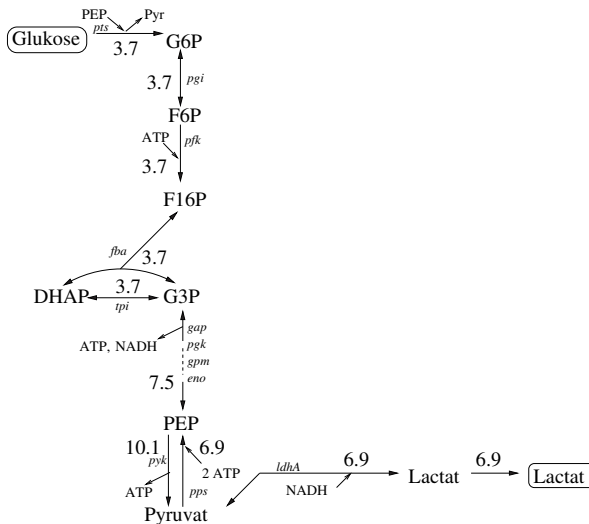


Abbildung 8.3: Differenz der Flussverteilungen („mit“ minus „ohne“ *ppsA*-Induktion). Die Beträge der Differenzen aller anderen Flüsse betragen maximal 0.4 mmol/(gDW h) und sind daher nicht von großer Bedeutung.

Es kann somit geschlossen werden, dass durch die Überexpression von *ppsA* ein gewünschtes Verhalten erzwingen werden konnte. Die Produktionsraten von Succinat und Lactat wurden im betrachteten Zeitraum signifikant erhöht, ohne die Wachstumsrate negativ zu beeinflussen. Die Analyse der intrazellulären Flussverteilung legt nahe, dass dies primär durch eine Erhöhung des glykolytischen Flusses erreicht wurde. Es konnte hier somit zum ersten mal gezeigt werden, dass ein erhöhter ATP-Verbrauch auch unter anaeroben Bedingungen zu einer gesteigerten Produktivität fermentativer Produkte führen kann. Somit konnte die mittels CASOP identifizierte Intervention zur Steigerung der Produktivität von Succinat und Lactat mit *E. coli* experimentell erfolgreich validiert werden. Für eine abschließende Beurteilung der implementierten Intervention muss jedoch noch das bisher ungeklärte Verhalten der Zellen nach acht Stunden aufgeklärt werden. Da dies induzierte als auch uninduzierte Zellen betrifft, ist es allerdings unwahrscheinlich, dass dieses Verhalten mit der Überexpression von *ppsA* in Verbindung steht.

9 Zusammenfassung

Folgende Schwerpunkte wurden in der vorliegenden Arbeit behandelt:

1. Die Entwicklung neuer Methoden zur Identifizierung von Interventionsstrategien in metabolischen Netzwerken.
2. Die Demonstration dieser Methoden anhand zu bestimmender Interventionsstrategien zur Produktion der industriell relevanten Metabolite Succinat, Lactat und Lycopin mit *E. coli*.
3. Die experimentelle Validierung ausgewählter Prädiktionen.

Ein Ziel des Metabolic Engineering ist es, durch gezielte Eingriffe die Stoffflüsse von Mikroorganismen oder Zellen derart umzulenken, dass ein gewünschtes Produkt vermehrt gebildet wird. Die molekularbiologischen Werkzeuge zur Umsetzung dieser Eingriffe sind Knockouts bzw. Überexpressionen von Genen. Mathematische Methoden sind für die modellbasierte Identifizierung geeigneter Interventionen von großer Bedeutung. Bisherige Ansätze verfolgen hauptsächlich das Ziel, die Produktausbeute unter wachstumsoptimalen Bedingungen zu optimieren und obligatorisch an die Biomassessynthese zu koppeln. Diese Kopplung kann bei den verwendeten Methoden unter suboptimalen Wachstumsbedingungen jedoch meist nicht garantiert werden. Durch die starke Fokussierung auf die optimale Ausbeute können andere wichtige Eigenschaften der Mutanten, wie z.B. die Produktivität, negativ beeinflusst werden.

Zwei neue Methoden, CASOP und constrained Minimal Cut Sets, wurden hier entwickelt. Mit diesen können Interventionen sowohl für das Strategieziel optimaler Produktausbeuten als auch für das Ziel einer hohen Produktivität vorgeschlagen werden. Während ausbeuteorientierte Methoden stringent sind, können bezüglich der Produktivität nur heuristische Ansätze verfolgt werden, da die kinetischen Eigenschaften des Netzwerkes nicht bekannt sind. Vielmehr erfolgt die Identifizierung der Interventionen unter Berücksichtigung der Netzwerkflexibilität, die als Approximation der Netzwerkkapazität und damit des potentiellen Stoffumsatzes interpretiert wird.

Die elementarmodenbasierte Heuristik CASOP liefert durch die Analyse produktivitätsbeeinflussender, struktureller Eigenschaften des Netzwerkes wichtige Informationen. Diese werden genutzt, um Interventionen für die (unter Umständen kombinierten) Strategieziele Ausbeute oder Produktivität zu identifizieren. Es ist theoretisch möglich, dass eine Reaktion in Abhängigkeit des gewählten Strategieziels entweder als Knockout- oder als Überexpressionskandidat identifiziert wird. Auf der einen Seite erschwert dies die

Entscheidung für eine bestimmte Intervention. Auf der anderen Seite wird die wertvolle Information gewonnen, dass die entsprechende Reaktion eine ambivalente Rolle für die Produktsynthese spielt. So birgt die Deletion aller suboptimalen Routen (mit potentiell hoher Kapazität) die Gefahr instabiler Stämme bzw. der Entwicklung von Stämmen mit niedriger Produktivität. CASOP ermöglicht darüber hinaus die Identifizierung von Überexpressionskandidaten, was nur wenige andere Methoden leisten. Zudem ist die Analyse der Bedeutung der Co-Faktoren im Zusammenhang mit der Produktsynthese besonders nützlich. Regulatorische oder technische Nebenbedingungen können einfach berücksichtigt werden, indem die Menge der aktiven Reaktionen oder EM entsprechend reduziert wird.

Die eingeführte Verallgemeinerung der MCS zu constrained MCS ermöglicht die flexible Formulierung komplexer Strategieziele, die mit den Ansätzen der linearen Optimierung nicht zugänglich sind. Die Definition erwünschter sowie unerwünschter Funktionalitäten kann bequem mittels geeigneter Mengen von Desired-EM und Target-EM erfolgen. Die große Flexibilität des Ansatzes wird durch die Reformulierung anderer Methoden als spezielle cMCS-Probleme belegt. Zudem kann mittels einer Mindestanzahl zu erhaltender EM die verbleibende Flexibilität des Systems kontrolliert werden. Ein entscheidender Vorteil des cMCS-Ansatzes gegenüber FBA-basierten Methoden ist die vollständige Enumeration aller äquivalenten Knockout-Strategien. Einige *in silico* Knockout-Kombinationen können experimentell nur schwer bis gar nicht realisierbar sein. Daher ist es sehr vorteilhaft, aus einer Menge gleichwertiger Strategien diejenige auswählen zu können, die in Bezug auf die praktische Umsetzbarkeit als die geeignetste erscheint. Aus der vollständigen Menge aller cMCS lassen sich zudem wichtige strukturelle Eigenschaften ableiten. So können unter anderem Knockouts identifiziert werden, die in jeder multiplen Knockout-Strategie vorkommen und somit essentiell für die Erreichung eines bestimmten Strategieziels sind. Durch einfache Ersetzungsregeln (basierend auf Gen-Reaktions-Assoziationen) können die Reaktionselementarmoden durch Elementarmoden der Gene substituiert werden. Dies ermöglicht die direkte Identifizierung von Gen-Knockout-Strategien.

Elementarmodenanalysen haben gegenüber anderen Ansätzen den Vorteil, dass sie eine unverzerrte Netzwerkanalyse durch die Betrachtung aller verfügbaren optimalen und suboptimalen Pathways erlauben. Es ist daher sehr erstrebenswert, diese Analysen auch in genomskaligen Netzwerken durchführen zu können. Der vorgestellte EFMSampler ist ein neuer Ansatz zur Berechnung von Elementarmoden in genomskaligen Netzwerken. An Stelle der vollständigen Menge der EM wird dabei eine repräsentative Teilmenge von EM generiert, die wesentliche Eigenschaften der vollständigen Menge hinreichend gut erfassen kann. Dadurch wird eine Vielzahl von elementarmodenbasierten Methoden für die Analyse genomskaliger Netzwerke zugänglich. Demonstriert wurde das in der vorliegenden Arbeit durch die Anwendung des EFMSampler in Kombination mit CASOP zur Identifizierung von Interventionen zur Produktion von Lactat, Succinat und Lycopin mit einem genomskaligen Netzwerk für *E. coli*. In einer anderen Studie wurde der EFMSampler in Kombination mit CASOP bereits erfolgreich zur Identifizierung von Knockouts zur Überproduktion von Aminosäuren mit *E. coli* genutzt (Pande et al., 2014).

Die Methode der cMCS wurde darüber hinaus inzwischen auch in genomskalige Netzwerke transferiert (von Kamp und Klamt, 2014). Mittels Dualitätsprinzipien können die kleinsten cMCS berechnet werden ohne zuvor die EM berechnen zu müssen.

In den Fallstudien wurden Succinat und Lactat als Vertreter von Produkten des Zentralstoffwechsels gewählt. Für die Steigerung ihrer Produktion in *E. coli* wurden durch CASOP viele Interventionen identifiziert, die experimentell bereits erfolgreich implementiert werden konnten. Dies betrifft unter anderem die Ausscheidungen der Nebenprodukte Acetat, Ethanol und Formiat als Knockout-Kandidaten. Weitere, weniger intuitive Interventionen, die in Übereinstimmung mit Ergebnissen aus der Literatur stehen, sind z.B. der Knockout des Zucker-PTS-Systems für die Produktion von Succinat oder der Fumarat-Reduktase für die Produktion von Lactat. Des Weiteren wurden die für die Succinat-Produktion bekannten Überexpressionskandidaten *aceAB* und *frd* durch CASOP identifiziert. Über die bekannten Interventionen hinaus wurden neue Interventionen vorgeschlagen.

Dabei ist insbesondere die gesteigerte ATP-Hydrolyse und die Überexpression von *ppsA* für die Produktion von Succinat und Lactat ein wenig intuitives Interventionsziel, das nur durch die systemweite Betrachtung des Netzwerks und des ATP-Haushalts unter verschiedenen Bedingungen verstanden werden kann. Für die Lactat-Produktion mit *E. coli* wurden bisher keine Überexpressionen beschrieben, so dass die hier identifizierten Kandidaten als Ausgangspunkt zukünftiger experimenteller Studien dienen können.

Mit dem Ansatz der constrained MCS wurden für Succinat und Lactat mehrere Tausend bzw. Hunderte verschiedene Knockout-Kombinationen identifiziert. Neben einzelnen Elementen (Reaktionen) identifizierter Knockout-Strategien wurden für beide Produkte prädierte komplexe Strategien bereits erfolgreich experimentell implementiert. Für die Produktion von Succinat konnte zudem gezeigt werden, dass viele Knockout-Kombinationen existieren, mit denen eine obligatorische CO₂-Fixierung erreicht werden kann.

Für Lycopin als Repräsentant eines Produkts des Sekundärstoffwechsels ist ebenfalls eine sehr gute Übereinstimmung mit bekannten Interventionen aus der Literatur festzustellen. Für aerobe Bedingungen wird in jeder der ersten drei CASOP-Iterationen mindestens ein bekannter Knockout identifiziert wie z.B. die Ausscheidung von Ethanol und Lactat sowie die Transhydrogenase. Der DXP-pathway als intuitives Interventionsziel wird von CASOP als bester Überexpressionskandidat prädiert.

Noch keine der mittels der cMCS identifizierten Strategien wurde bisher vollständig experimentell umgesetzt. Es existiert bisher allerdings auch noch keine Mutante, die Lycopin mit den für die Berechnung der cMCS verwendeten Anforderungen produzieren kann. Da einige Elemente der prädierten Strategien wie z.B. *aceE* oder *pgi* bereits experimentell implementiert wurden, können die hier vorgeschlagenen Knockout-Kombinationen zur Optimierung der existierenden Produktionsstämme herangezogen werden. Für anaerobe Bedingungen gibt es bisher keine Studien, die die Produktion von Lycopin mit *E. coli* beschreiben, so dass die experimentelle Überprüfung der identifizierten Interventionen noch aussteht.

Die erhöhte ATP-Hydrolyse als klares Interventionsziel zur Steigerung der Produktion von Succinat und bzw. oder Lactat unter anaeroben Bedingungen ist das unintuitiv-

ste der vorgeschlagenen Interventionsziele. Daher wurde es zur Validierung durch eigene Experimente ausgewählt. Dafür wurde ein induzierbarer ATP futile cycle durch die Überexpression von *ppsA* implementiert. Im Gegensatz zu Ergebnissen aus der Literatur für aerobe Bedingungen konnte unter anaeroben Bedingungen und dem verwendeten experimentellen Aufbau kein Einfluss der *ppsA*-Überexpression auf die Wachstumsrate beobachtet werden. Dies deutet darauf hin, dass *E. coli* unter diesen Bedingungen nicht durch ATP limitiert ist bzw. den Bedarf an ATP durch eine gesteigerte Substrataufnahme ausgleichen kann. Die Aktivierung des futile cycles durch IPTG führte zu einer gesteigerten Glukoseaufnahme sowie zu einer erhöhten Succinat- und Lactat-Ausscheidungsrate von ca. 36%, womit die Prädiktion erfolgreich validiert werden konnte. Der Ansatz der erhöhten ATP-Hydrolyse zur Steigerung der Produktivität kann grundsätzlich auch für andere Produkte und Mikroorganismen verwendet werden. Dies gilt unter der Voraussetzung, dass bei der Produktsynthese ein ATP-Überschuss generiert werden kann und alternative ATP-Quellen eliminiert wurden. Eine (künstliche) Reduktion der ATP-Verfügbarkeit kann dann nur durch eine erhöhte Produktsynthese ausgeglichen werden. Unter der Annahme, dass der Organismus optimale Wachstumsraten anstrebt (die viel ATP benötigen), muss somit auch die Rate der Produktsynthese optimiert werden.

In Zukunft sind weitere Experimente notwendig, um die Ursachen der reduzierten Wachstumsrate nach acht Stunden aufzuklären bzw. die Stabilität des Stammes zu verbessern.

Eine weitere experimentelle Validierung einer mittels CASOP identifizierten Intervention wurde durch eine kürzlich vorgestellte Studie vorgenommen (Yang et al., 2013). Dabei wurden verschiedene carboxylierende Reaktionen auf ihren Einfluss bezüglich der Succinat-Produktivität untersucht. Die Pyruvat-Carboxylase wurde von CASOP als geeignetster Überexpressionskandidat identifiziert und experimentell implementiert. Der resultierende *E. coli*-Stamm zeigt die höchste bisher erreichte Succinat-Produktivität unter aeroben Bedingungen.

Für die Beurteilung der Relevanz beobachteter Unterschiede metabolischer Raten zwischen verschiedenen Umweltbedingungen wird deren statistische Signifikanz untersucht. Von besonderer Bedeutung sind dabei korrekt bestimmte Varianzen der spezifischen Raten. Neben der Genauigkeit der Konzentrationsmessungen wird die Gesamtvarianz durch die biologische Varianz beeinflusst, die zwischen unabhängigen Experimenten auftritt. Um falsch-positive Resultate zu vermeiden, muss diese biologische Varianz für eine unverzerrte Analyse der Beobachtungen berücksichtigt werden. Hier wurde ein generischer Ansatz aufgezeigt, wie diese Analysen mit grundlegenden statistischen Verfahren behandelt werden können.

Mit Hilfe der industriellen Biotechnologie können eine Vielzahl verschiedener Substanzen durch den Einsatz geeigneter, evtl. modifizierter Mikroorganismen produziert werden. Zukünftige Herausforderungen liegen unter Anderem in der Verwendung erneuerbarer und kostengünstiger Ausgangsmaterialien sowie in der Steigerung der Effizienz der eingesetzten Mikroorganismen. Die hier vorgestellten Methoden können einen Beitrag bei der Konstruktion der eingesetzten Produktionsstämme leisten und damit die Entwicklung ökologisch nachhaltiger und ökonomisch kompetitiver Prozesse unterstützen.

A Anhang

A.1 Zentralstoffwechselmodell für *Escherichia Coli*

Einige Co-Faktoren wie z.B. NAD, NADP, ADP und CoA sind nicht explizit in den Stöchiometrien der Reaktionen berücksichtigt, da sie in Erhaltungsrelationen mit den zugehörigen komplementären Metaboliten (NADH, NADPH, ATP, ...) auftreten. Gennamen sind für die Reaktionen zentraler Stoffwechselwege gegeben und werden auch als Abkürzung der Reaktionsnamen im Haupttext verwendet. Existieren mehrere Gene für eine Reaktion, so wird nur ein repräsentatives Gen angegeben (Isoenzyme werden vernachlässigt).

Stöchiometrisches Modell des *E. coli* Zentralstoffwechsels (Katabolismus).

Pfad	Reaktionsname	Reaktion	Gen
Aufnahme	O2_up	$O2(\text{external}) \Rightarrow O2$	
	CO2_up	$CO2(\text{external}) \Rightarrow CO2$	
	Glc_up	$\Rightarrow \text{Glc}$	
	Glc_PTS_up	$\text{Glucose}(\text{external}) + \text{PEP} \Rightarrow \text{G6P} + \text{Pyr}$	<i>ptsG</i>
	Glc_ATP_up	$\text{Glucose}(\text{external}) + \text{ATP} + \text{H}_{\text{ex}} \Rightarrow \text{G6P}$	<i>glk</i>
	N_up	$\text{N}(\text{external}) \Rightarrow \text{N}$	
Glykolyse	S_up	$\text{S}(\text{external}) + 4 \text{ ATP} + 4 \text{ NADPH} \Rightarrow \text{S}$	
	G6P::F6P	$\text{G6P} \Leftrightarrow \text{F6P}$	<i>pgi</i>
	F6P::F16P	$\text{F6P} + \text{ATP} \Rightarrow \text{F16P}$	<i>pfkA</i>
	F16P::T3P	$\text{F16P} \Leftrightarrow \text{DHAP} + \text{G3P}$	<i>fbaA</i>
	DHAP::G3P	$\text{DHAP} \Leftrightarrow \text{G3P}$	<i>tpiA</i>
	G3P::DPG	$\text{G3P} \Leftrightarrow \text{DPG} + \text{NADH}$	<i>gapA</i>
	DPG::3PG	$\text{DPG} \Leftrightarrow 3\text{PG} + \text{ATP}$	<i>pgk</i>
	3PG::2PG	$3\text{PG} \Leftrightarrow 2\text{PG}$	<i>gpmA</i>
	2PG::PEP	$2\text{PG} \Leftrightarrow \text{PEP}$	<i>eno</i>
	PEP::PYR	$\text{PEP} \Rightarrow \text{Pyr} + \text{ATP}$	<i>pykA</i>
	Pyr::PEP	$\text{Pyr} + 2 \text{ ATP} \Rightarrow \text{PEP}$	<i>pps</i>
Pentose-Phosphat-Weg	G6P::PGLac	$\text{G6P} \Leftrightarrow \text{PGLac} + \text{NADPH}$	<i>zwf</i>
	PGLac::PGLuc	$\text{PGLac} \Rightarrow \text{PGLuc}$	<i>pgl</i>
	PGLuc::R15P	$\text{PGLuc} \Rightarrow \text{R15P} + \text{NADPH} + \text{CO2}$	<i>gnd</i>
	R15P::X5P	$\text{R15P} \Leftrightarrow \text{X5P}$	<i>rpe</i>
	R15P::R5P	$\text{R15P} \Leftrightarrow \text{R5P}$	<i>rpiA</i>
	Transket1	$\text{R5P} + \text{X5P} \Leftrightarrow \text{G3P} + \text{S7P}$	<i>tktA</i>
	Transaldo	$\text{G3P} + \text{S7P} \Leftrightarrow \text{F6P} + \text{E4P}$	<i>talA</i>
Entner-Doudoroff-Weg	Transket2	$\text{E4P} + \text{X5P} \Leftrightarrow \text{F6P} + \text{G3P}$	<i>tktA</i>
	2KD6PG::PyrG3P	$2\text{KD6PG} \Rightarrow \text{G3P} + \text{Pyr}$	<i>eda</i>
	PGLuc::2KD6PG	$\text{PGLuc} \Rightarrow 2\text{KD6PG}$	<i>edd</i>

Fortsetzung des stöchiometrischen Modells des *E. coli* Zentralstoffwechsels (Katabolismus).

Pfad	Reaktionsname	Reaktion	Gen
Anaplerotische Reaktionen	Mal::PyrNADH	$\text{Mal} \Rightarrow \text{Pyr} + \text{NADH} + \text{CO}_2$	<i>maeA</i>
	OxA::PEP	$\text{OxA} + \text{ATP} \Rightarrow \text{PEP} + \text{CO}_2$	<i>pck</i>
	PEP::OxA	$\text{PEP} + \text{CO}_2 \Rightarrow \text{OxA}$	<i>ppc</i>
Citrat-Zyklus	PYR::AcCoA	$\text{Pyr} \Rightarrow \text{AcCoA} + \text{NADH} + \text{CO}_2$	<i>aceE</i>
	AcCoA::Cit	$\text{AcCoA} + \text{OxA} \Rightarrow \text{Cit}$	<i>gltA</i>
	Cit::ICit	$\text{Cit} \Leftrightarrow \text{ICit}$	<i>acnA</i>
	ICit::alKG	$\text{ICit} \Leftrightarrow \text{alKG} + \text{NADPH} + \text{CO}_2$	<i>icd</i>
	alKG::SuccCoA	$\text{alKG} \Rightarrow \text{SuccCoA} + \text{NADH} + \text{CO}_2$	<i>sucAB</i>
	SuccCoA::Succ	$\text{SuccCoA} \Leftrightarrow \text{Succ} + \text{ATP}$	<i>sucCD</i>
	Succ::Fum	$\text{Succ} \Rightarrow \text{Fum} + \text{QuiH}_2$	<i>sdhAB</i>
	Fum::Succ	$\text{Fum} + \text{QuiH}_2 \Rightarrow \text{Succ}$	<i>frdA</i>
	Fum::Mal	$\text{Fum} \Leftrightarrow \text{Mal}$	<i>fumA</i>
Glyoxylat-Weg	Mal::OxA_NADH	$\text{Mal} \Leftrightarrow \text{OxA} + \text{NADH}$	<i>mdh</i>
	ICit::Glyox	$\text{ICit} \Rightarrow \text{Succ} + \text{Glyox}$	<i>aceA</i>
Fermentative Reaktionen	Glyox::Mal	$\text{AcCoA} + \text{Glyox} \Rightarrow \text{Mal}$	<i>aceB</i>
	AcCoA::Adh	$\text{AcCoA} + \text{NADH} \Leftrightarrow \text{Adh}$	<i>adhE</i>
	Adh::Eth	$\text{NADH} + \text{Adh} \Leftrightarrow \text{Eth}$	<i>adhE</i>
	AcCoA::AcP	$\text{AcCoA} \Leftrightarrow \text{AcP}$	<i>pta</i>
	AcP::Ac	$\text{AcP} \Leftrightarrow \text{ATP} + \text{Ac}$	<i>ackA</i>
	Pyr::Form	$\text{Pyr} \Rightarrow \text{AcCoA} + \text{Form}$	<i>pflB</i>
	Form_ex	$\text{Form} \Rightarrow \text{Form}(\text{external})$	<i>focA</i>
	Pyr::Lac	$\text{Pyr} + \text{NADH} \Leftrightarrow \text{Lac}$	<i>ldhA</i>
	Pyr::Ac	$\text{Pyr} \Rightarrow \text{QuiH}_2 + \text{CO}_2 + \text{Ac}$	<i>poxB</i>
Elektronen-Transport-Kette	DHAP::Lac	$\text{DHAP} \Rightarrow \text{Lac}$	<i>mgsA</i>
	NADHDehydro	$\text{NADH} \Leftrightarrow \text{QuiH}_2 + 2 \text{H}_{\text{ex}}$	<i>nuoA</i>
	Oxidase	$\text{QuiH}_2 + 0.5 \text{O}_2 \Rightarrow 2 \text{H}_{\text{ex}}$	<i>cyoA</i>
	TransHydro	$\text{NADH} + \text{H}_{\text{ex}} \Leftrightarrow \text{NADPH}$	<i>pntA</i>
	ATPSynth	$3 \text{H}_{\text{ex}} \Leftrightarrow \text{ATP}$	<i>atp</i>
DXP-Pathway	ATPdrain	$\text{ATP} \Rightarrow (\text{maintenance})$	
	PyrG3P::DXP	$\text{G3P} + \text{Pyr} \Rightarrow \text{CO}_2 + \text{DXP}$	<i>dxs</i>
	DXP::MEP	$\text{NADPH} + \text{DXP} \Rightarrow \text{MEP}$	<i>dxr</i>
	MEP::HMBDP	$3 \text{ATP} + \text{NADPH} + \text{MEP} \Rightarrow \text{HMBDP}$	<i>ispDEFG</i>
	HMBDP::DMAPP	$\text{NADPH} + \text{HMBDP} \Rightarrow \text{DMAPP}$	<i>ispH</i>
	HMBDP::IPP	$\text{NADPH} + \text{HMBDP} \Rightarrow \text{IPP}$	<i>ispH</i>
Lycopin-Synthese	DMAPP::IPP	$\text{DMAPP} \Leftrightarrow \text{IPP}$	<i>idi</i>
	IPP::GPP	$\text{DMAPP} + \text{IPP} \Rightarrow \text{GPP}$	<i>ispA</i>
	GPP::FPP	$\text{IPP} + \text{GPP} \Rightarrow \text{FPP}$	<i>ispA</i>
	FPP::GGPP	$\text{IPP} + \text{FPP} \Rightarrow \text{GGPP}$	<i>crtE</i>
	GGPP::Phytoene	$2 \text{GGPP} \Rightarrow \text{Phytoene}$	<i>crtB</i>
Ausscheidung	Phytoene::Lycopene	$\text{Phytoene} \Rightarrow \text{Lycopene}$	<i>crtI</i>
	Eth_ex	$\text{Eth} \Rightarrow \text{Eth}(\text{external})$	
	Lac_ex	$\text{Lac} \Rightarrow \text{Lac}(\text{external})$	
	Ac_ex	$\text{Ac} \Rightarrow \text{Ac}(\text{external})$	
	Succ_Ex	$\text{Succ} \Rightarrow \text{Succ}(\text{external})$	
	Lycopene_ex	$\text{Lycopene} \Rightarrow \text{Lyco}_{\text{ex}}$	
	CO2_ex	$\text{CO}_2 \Rightarrow \text{CO}_2(\text{external})$	

Fortsetzung des stöchiometrischen Modells des *E. coli* Zentralstoffwechsels (Anabolismus).

Reaktionsname	Reaktion
DHAP::Glyc3P	DHAP + NADH $\Rightarrow$ Glyc3P
Chor_Synth	2 PEP + E4P + ATP + NADPH $\Rightarrow$ Chor
PRPP_Synth	R5P + 2 ATP $\Rightarrow$ PRPP
MTHF_Synth2	Gly $\Leftrightarrow$ MTHF + CO2 + NADH + N
Ala_Synth	Pyr + Glu $\Rightarrow$ alKG + Ala
Val_Synth	2 Pyr + NADPH + Glu $\Rightarrow$ alKG + CO2 + Val
Leu_Synth	2 Pyr + AcCoA + NADPH + Glu $\Rightarrow$ alKG + NADH + 2 CO2 + Leu
Asn_Synth_1	2 ATP + N + Asp $\Rightarrow$ Asn
Asp_synth	OxA + Glu $\Rightarrow$ alKG + Asp
Asp::AspSAld	ATP + NADPH + Asp $\Rightarrow$ AspSAld
AspSAld::HSer	NADPH + AspSAld $\Rightarrow$ HSer
Lys_Synth	di_am_pim $\Rightarrow$ CO2 + Lys
Met_Synth	SuccCoA + MTHF + HSer + Cys $\Rightarrow$ Pyr + Succ + N + Met
Thr_Synth	ATP + HSer $\Rightarrow$ Thr
Ile_Synth	Pyr + NADPH + Glu + Thr $\Rightarrow$ alKG + CO2 + N + Ile
His_Synth	ATP + PRPP + Gln $\Rightarrow$ alKG + 2 NADH + His
Glu_synth	alKG + NADPH + N $\Rightarrow$ Glu
Gln_Synth	ATP + N + Glu $\Rightarrow$ Gln
Pro_Synth	ATP + 2 NADPH + Glu $\Rightarrow$ Pro
Arg_Synth	AcCoA + 4 ATP + NADPH + CO2 + N + Asp + 2 Glu $\Rightarrow$ alKG + Fum + Ac + Arg
Trp_Synth	Chor + PRPP + Gln + Ser $\Rightarrow$ G3P + Pyr + CO2 + Glu + Trp
Tyr_Synth	Chor + Glu $\Rightarrow$ alKG + NADH + CO2 + Tyr
Phe_Synth	Chor + Glu $\Rightarrow$ alKG + CO2 + Phe
Ser_Synth	3PG + Glu $\Rightarrow$ alKG + NADH + Ser
Gly_Synth	Ser $\Rightarrow$ MTHF + Gly
Cys_Synth	AcCoA + S + Ser $\Rightarrow$ Ac + Cys
rATP_Synth	5 ATP + CO2 + PRPP + 2 MTHF + 2 Asp + Gly + 2 Gln $\Rightarrow$ 2 Fum + NADPH + 2 Glu + rATP
rGTP_Synth	6 ATP + CO2 + PRPP + 2 MTHF + Asp + Gly + 3 Gln $\Rightarrow$ Fum + NADH + NADPH + 3 Glu + rGTP
rCTP_Synth	ATP + Gln + rUTP $\Rightarrow$ Glu + rCTP
rUTP_Synth	4 ATP + N + PRPP + Asp $\Rightarrow$ NADH + rUTP
dATP_Synth	NADPH + rATP $\Rightarrow$ dATP
dGTP_Synth	NADPH + rGTP $\Rightarrow$ dGTP
dCTP_Synth	NADPH + rCTP $\Rightarrow$ dCTP
dTTP_Synth	2 NADPH + MTHF + rUTP $\Rightarrow$ dTTP
mit_FS_Synth	8.24 AcCoA + 7.24 ATP + 13.91 NADPH $\Rightarrow$ mit_FS
UDPGlc_Synth	G6P + ATP $\Rightarrow$ UDPGlc
CDPEth_Synth	3PG + 3 ATP + NADPH + N $\Rightarrow$ NADH + CDPEth
OH_myrc_ac_Synth	7 AcCoA + 6 ATP + 11 NADPH $\Rightarrow$ OH_myrc_ac
C14_0_FS_Synth	7 AcCoA + 6 ATP + 12 NADPH $\Rightarrow$ C14_0_FS
CMP_KDO_Synth	PEP + R5P + 2 ATP $\Rightarrow$ CMP_KDO
NDPHep_Synth	1.5 G6P + ATP $\Rightarrow$ 4 NADPH + NDPHep
TDPGlc_Synth	F6P + 2 ATP + N $\Rightarrow$ TDPGlc
UDP_NAG_Synth	F6P + AcCoA + ATP + Gln $\Rightarrow$ Glu + UDP_NAG
UDP_NAM_Synth	PEP + NADPH + UDP_NAG $\Rightarrow$ UDP_NAM
di_am_pim_Synth	Pyr + SuccCoA + NADPH + AspSAld + Glu $\Rightarrow$ alKG + Succ + di_am_pim
ADPGlc_Synth	G6P + ATP $\Rightarrow$ ADPGlc

Stöchiometrie der Biomassesynthese für 1 Gramm Biotrockenmasse von *E. coli*.

Reaktionsname	Stöchiometrie
μ :	0.14176 Glyc3P + 26.2949 ATP + 0.60097 Ala + 0.10124 Cys + 0.26647 Asp + 0.30747 Glu + 0.2048 Phe + 0.67725 Gly + 0.10473 His + 0.32116 Ile + 0.37935 Lys + 0.49804 Leu + 0.16989 Met + 0.26647 Asn + 0.24436 Pro + 0.29091 Gln + 0.32698 Arg + 0.38031 Ser + 0.28044 Thr + 0.46778 Val + 0.062835 Trp + 0.15244 Tyr + 0.1489 rATP + 0.18319 rGTP + 0.11366 rCTP + 0.12273 rUTP + 0.023904 dATP + 0.024582 dGTP + 0.024582 dCTP + 0.023904 dTTP + 0.28352 av_FA + 0.0069264 UDPGlc + 0.010368 CDPEth + 0.010368 OH_myrr_ac + 0.010368 C14_0_FS + 0.010368 CMP_KDO + 0.010368 NDPHep + 0.0069264 TDPGlc + 0.01656 UDP_NAG + 0.01656 UDP_NAM + 0.01656 di_am_pim + 0.0924 ADPGlc $\Rightarrow$ Biomasse

Metabolite des *E. coli* Zentralstoffwechselmodells.

Bezeichner	Vollständiger Name	Bezeichner	Vollständiger Name
2KD6PG	2-Keto-3-Deoxy-6-Gluconat	HSer	Homoserin
2PG	2-Phosphoglycerat	ICit	Isocitrat
3PG	3-Phosphoglycerat	Ile	Isoleucin
Ac	Acetat	IMP	Inosinmonophosphat
AcCoA	Acetyl-CoA	IPP	Isopentenyl Diphosphat
AcLac	Acetolactat	Lac	Lactat
AcP	Acetyl-P	Leu	Leucin
Adh	Acetaldehyd	Lip	Lipide
ADPGlc	ADP-Glukose	LPS	Lipopolysaccharide
Ala	Alanin	Lycopene	Lycopin
AKG	α -Ketoglutarat	Lys	Lysin
Arg	Arginin	Mal	Malat
Asn	Asparagin	MEP	2C-Methyl-D-Erythriol
Asp	Aspartat	Met	Methionin
AspSAld	Aspartat-4-Semialdehyd	MTHF	Methylen-Tetrahydrofolat
ATP	Adenosintriphosphat	N	Ammonium
av_g_FS	durchschnittliche Fettsäure	NADH	NADH
C14_0_FS	14:0 Fettsäure	NADPH	NADPH
CDPEth	CDP-Ethanolamin	NDPHep	NDP-Heptose
Chor	Chorismat	O2	Sauerstoff
Cit	Citrat	OH_myrr_ac	OH myristische Säure
CMP_KDO	CMP-KDO	OAA	Oxalacetat
CO2	Kohlenstoffdioxid	PEP	Phosphoenolpyruvat
Cys	Cystein	PepGly	Peptidoglykane
dATP	ATP für DNA Synthese	PGLac	6-Phospho-Gluconolacton
dCTP	CTP für DNA Synthese	PGLuc	6-Phospho-Gluconat
dGTP	GTP für DNA Synthese	Phe	Phenylalanin
DHAP	Dihydroxyaceton-Phosphat	Phytoene	Phytoen

Fortsetzung der Metabolite des *E. coli* Zentralstoffwechselmodells.

Bezeichner	Vollständiger Name	Bezeichner	Vollständiger Name
diampim	Diaminopimelat	PolyPhos	Polyphosphate
DMAPP	Dimethylallyl Diphosphat	Pro	Prolin
DNA	Desoxyribonukleinsäure	PRPP	5-Phosphoribosyl-1-Pyrophosphat
DPG	Diphosphoglycerat	Pyr	Pyruvat
dTTP	TTP für DNA Synthese	QH2	Ubiquinon
DXP	1-Deoxy-D-Xylulose-5-Phosphat	R5P	Ribose-5-Phosphat
E4P	Erythrose-4-Phosphat	rATP	ATP für RNA Synthese
Eth	Ethanol	rCTP	CTP für RNA Synthese
F16P	Fructose-1,6-Bisphosphat	rGTP	GTP für RNA Synthese
F6P	Fructose-6-Phosphat	RL5P	Ribulose-5-Phosphat
Form	Format	RNA	Ribonukleinsäure
FPP	Farnesyl Diphosphat	rUTP	UTP für RNA Synthese
Fum	Fumarat	S	Schwefel (SO4)
G3P	Glyceraldehyde-3-Phosphat	S7P	Seduheptulose-7-Phosphat
G6P	Glucose-6-Phosphat	Ser	Serin
GGPP	Geranylgeranyl Diphosphat	Shiki	Shikimat
Gln	Glutamin	Succ	Succinat
Glu	Glutamat	SuccCoA	Succinyl-Coenzym-A
Gly	Glycin	TDPGlc	TDP-Glucosamin
Glyc	Glycogen	Thr	Threonin
Glyc3P	Glycerol-3-Phosphat	Trp	Tryptophan
Glyox	Glyoxylat	Tyr	Tyrosin
GPP	Geranyl Diphosphat	UDP_NAG	UDP-Acetylglucosamin
H_ex	periplasmatische Protonen	UDP_NAM	UDP-N-Acetylmuramische Säure
His	Histidin	UDPGlc	UDP-Glukose
HMBDP	1-Hydroxy-2-Methyl- Butenyl-4-Diphosphat	Val	Valin
		X5P	Xylulose-5-Phosphat

A.2 p -Werte der statistischen Tests mit unvollständigen Daten

p -Werte der statistischen Tests mit unvollständigen Daten. Für alle Kombination von zwei biologisch unabhängigen Experimenten wird ein z -Test mit jeweils zwei verschiedenen Varianzen bzw. Standardabweichungen durchgeführt (der Methodenvarianz σ_M und der spezifischen Gesamtvarianz). *: signifikant nach Bonferroni-Korrektur ($p < 8.3E - 03$ für Wave (W) vs. STR, $p < 1.7E - 02$ für STR vs. STR)

	Glukose		Lactat		Ammonium		Glutamin		Glutamat		μ	
	σ_M	$\sigma_{4,2\%}$	σ_M	$\sigma_{16,5\%}$	σ_M	$\sigma_{16,9\%}$	σ_M	$\sigma_{19,4\%}$	σ_M	$\sigma_{73,2\%}$	σ_M	$\sigma_{10,7\%}$
W vs. STR1,2	2,7E-04*	1,1E-03*	7,9E-04*	0,31	1,3E-03*	5,2E-02	0,66	0,10	0,12	0,66	1,3E-06*	9,4E-02
W vs. STR1,3	1,3E-02	3,4E-02	4,1E-02	0,54	4,5E-03*	0,11	0,94	0,83	5,2E-02	0,63	0,92	0,97
W vs. STR1,4	6,7E-04*	5,4E-03*	0,83	0,95	0,29	0,55	0,63	0,15	7,9E-03*	0,57	1,4E-02	0,43
W vs. STR2,3	8,8E-03	1,7E-02	3,3E-04*	0,27	0,10	0,28	0,98	0,94	3,9E-04*	0,15	0,24	0,69
W vs. STR2,4	4,7E-04*	2,3E-03*	4,3E-02	0,54	0,95	0,96	0,64	0,11	8,9E-04*	0,23	0,24	0,69
W vs. STR3,4	2,1E-02	5,6E-02	5,3E-10*	6,2E-02	0,35	0,57	0,93	0,82	7,4E-04*	0,27	5,1E-04*	0,26
STR1,2 vs. STR 3,4	0,14	0,16	0,00*	5,1E-03*	5,1E-05*	1,4E-02*	0,62	6,3E-02	3,1E-05*	0,18	4,4E-16*	5,8E-03*
STR1,3 vs. STR 2,4	0,33	0,35	0,97	0,99	6,5E-03*	0,10	0,63	6,9E-02	2,7E-06*	0,15	0,21	0,67
STR1,4 vs. STR 2,3	0,67	0,68	7,1E-04*	0,29	0,43	0,62	0,69	0,13	1,6E-08*	0,10	5,2E-04*	0,23

A.3 Experimentelle Daten

Tabelle A.1: Konzentrationsdaten anaerobe Fermentation mit Induktion.

Zeit	Experiment 1a				Experiment 1b				Experiment 1a				Experiment 1b			
	OD	Glc	Succ	Lac	OD	Glc	Succ	Lac	-IPTG1	Glc	Succ	Lac	-IPTG2	Glc	Succ	Lac
0.0	0.20	4.00	0.00	0.00	0.20	4.00	0.00	0.00	0.20	4.00	0.00	0.00	0.20	4.00	0.00	0.00
1.4	0.25	3.96	13.23	89.83	0.25	3.83	7.44	92.51	0.26	3.91	3.31	93.18	0.28	3.88	6.62	87.82
1.5	0.25	3.96	13.23	89.83	0.25	3.83	7.44	92.51	0.26	3.91	3.31	93.18	0.28	3.88	6.62	87.82
2.4	0.33	3.68	5.79	172.25	0.34	3.46	43.93	189.97	0.32	3.62	-	154.58	0.31	3.49	-	156.37
3.1	0.36	3.54	-	269.42	0.36	3.34	-	295.29	0.34	3.52	-	228.40	0.35	3.26	-	225.51
3.8	0.41	3.25	51.27	419.58	0.41	3.15	39.94	456.83	0.39	3.41	40.70	306.01	0.39	3.35	42.39	325.66
4.4	0.46	3.16	64.50	528.74	0.48	2.96	-	549.23	0.43	3.34	-	350.81	0.42	3.25	-	377.36
4.9	0.50	3.06	67.81	636.00	0.51	2.92	67.10	672.00	0.46	3.34	48.95	418.95	0.46	3.12	57.61	451.02
5.6	0.55	2.88	92.62	742.00	0.58	2.65	-	832.25	0.50	3.21	50.91	511.07	0.50	3.03	38.66	560.86
6.1	0.57	2.73	92.24	1001.87	0.60	2.55	91.14	1006.19	0.54	3.09	55.89	661.75	0.53	2.88	74.03	660.63
6.8	0.62	2.51	113.06	1162.07	0.63	2.40	99.13	1173.52	0.58	3.02	58.50	788.69	0.58	2.82	76.84	829.32
7.4	0.66	2.36	124.58	1326.73	0.69	2.21	113.65	1351.67	0.64	2.84	60.03	922.47	0.65	2.66	74.31	979.18
8.1	0.73	2.07	134.18	1574.86	0.72	1.95	107.71	1587.03	0.65	2.70	64.60	1097.89	0.68	2.46	80.90	1210.61

Zeit	Experiment 2a				Experiment 2b				Experiment 2a				Experiment 2b			
	OD	Glc	Succ	Lac	OD	Glc	Succ	Lac	OD	Glc	Succ	Lac	OD	Glc	Succ	Lac
0.0	0.20	4.00	0.00	0.00	0.20	4.00	0.00	0.00	0.20	4.00	0.00	0.00	0.20	4.00	0.00	0.00
1.4	0.25	3.77	-	60.57	0.25	3.65	-	66.55	0.25	3.83	-	76.35	0.24	3.71	-	71.08
2.4	0.33	3.68	5.79	172.25	0.34	3.46	43.93	189.97	0.32	3.62	-	154.58	0.31	3.49	-	156.37
3.1	0.36	3.54	-	269.42	0.36	3.34	-	295.29	0.34	3.52	-	228.40	0.35	3.26	-	225.51
3.8	0.41	3.25	51.27	419.58	0.41	3.15	39.94	456.83	0.39	3.41	40.70	306.01	0.39	3.35	42.39	325.66
4.4	0.46	3.16	64.50	528.74	0.48	2.96	-	549.23	0.43	3.34	-	350.81	0.42	3.25	-	377.36
4.9	0.50	3.06	67.81	636.00	0.51	2.92	67.10	672.00	0.46	3.34	48.95	418.95	0.46	3.12	57.61	451.02
5.6	0.55	2.88	92.62	742.00	0.58	2.65	-	832.25	0.50	3.21	50.91	511.07	0.50	3.03	38.66	560.86
6.1	0.57	2.73	92.24	1001.87	0.60	2.55	91.14	1006.19	0.54	3.09	55.89	661.75	0.53	2.88	74.03	660.63
6.8	0.62	2.51	113.06	1162.07	0.63	2.40	99.13	1173.52	0.58	3.02	58.50	788.69	0.58	2.82	76.84	829.32
7.4	0.66	2.36	124.58	1326.73	0.69	2.21	113.65	1351.67	0.64	2.84	60.03	922.47	0.65	2.66	74.31	979.18
8.1	0.73	2.07	134.18	1574.86	0.72	1.95	107.71	1587.03	0.65	2.70	64.60	1097.89	0.68	2.46	80.90	1210.61

Literaturverzeichnis

- Acuña, V., Marchetti-Spaccamela, A., Sagot, M.-F., und Stougie, L. A note on the complexity of finding and enumerating elementary modes. *Biosystems*, 99(3):210–214, 2010.
- Adrio, J. L. und Demain, A. L. Genetic improvement of processes yielding microbial products. *FEMS Microbiol Rev*, 30(2):187–214, 2006.
- Adrio, J.-L. und Demain, A. L. Recombinant organisms for production of industrial products. *Bioeng Bugs*, 1(2):116–131, 2010.
- Almaas, E., Kovács, B., Vicsek, T., Oltvai, Z. N., und Barabási, A.-L. Global organization of metabolic fluxes in the bacterium *Escherichia coli*. *Nature*, 427(6977):839–843, 2004.
- Alper, H. und Stephanopoulos, G. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metab Eng*, 9(3):258–267, 2007. 095.
- Alper, H. und Stephanopoulos, G. Uncovering the gene knockout landscape for improved lycopene production in *E. coli*. *Appl Microbiol Biotechnol*, 2008. 091.
- Alper, H., Fischer, C., Nevoigt, E., und Stephanopoulos, G. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A*, 102(36):12678–12683, 2005a.
- Alper, H., Jin, Y.-S., Moxley, J. F., und Stephanopoulos, G. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metab Eng*, 7(3):155–164, 2005b. 093.
- Alper, H., Miyaoku, K., und Stephanopoulos, G. Construction of lycopene-overproducing *E. coli* strains by combining systematic and combinatorial gene knockout targets. *Nat Biotechnol*, 23(5):612–616, 2005c. 094.
- Alper, H., Moxley, J., Nevoigt, E., Fink, G. R., und Stephanopoulos, G. Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science*, 314(5805):1565–1568, 2006.
- Arriaga, E. A. Determining biological noise via single cell analysis. *Anal Bioanal Chem*, 393(1):73–80, 2009.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., und Mori, H. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol*, 2:2006.0008, 2006.

- Báez-Viveros, J. L., Flores, N., Juárez, K., na, P. C.-E., Bolivar, F., und Gosset, G. Metabolic transcription analysis of engineered *Escherichia coli* strains that overproduce L-phenylalanine. *Microb Cell Fact*, 6:30, 2007. 171.
- Ballerstein, K., von Kamp, A., Klamt, S., und Haus, U.-U. Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*, 28(3):381–387, 2012.
- Beauprez, J. J., De Mey, M., und Soetaert, W. K. Microbial succinic acid production: natural versus metabolic engineered producers. *Process Biochemistry*, 45(7):1103–1114, 2010.
- Bechthold, I., Bretz, K., Kabasci, S., Kopitzky, R., und Springer, A. Succinic acid: a new platform chemical for biobased polymers from renewable resources. *Chemical engineering & technology*, 31(5):647–654, 2008.
- Becker, S. A. und Palsson, B. Ø. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*, 4(5):e1000082, 2008. 259.
- Behre, J., Wilhelm, T., von Kamp, A., Ruppin, E., und Schuster, S. Structural robustness of metabolic networks with respect to multiple knockouts. *J Theor Biol*, 252(3):433–441, 2008. 162.
- Berge, C. *Hypergraphs. Combinatorics of finite sets*. North-Holland, Amsterdam, 1989.
- Bertsimas, D. und Tsitsiklis, J. *Introduction to linear optimization*. Athena Scientific, Belmont, 1997.
- Biot-Pelletier, D. und Martin, V. J. J. Evolutionary engineering by genome shuffling. *Appl Microbiol Biotechnol*, 2014.
- Blazier, A. S. und Papin, J. A. Integration of expression data in genome-scale metabolic network reconstructions. *Front Physiol*, 3:299, 2012.
- Bonarius, H. P., Houtman, J. H., Schmid, G., de Gooijer, C. D., und Tramper, J. Error analysis of metabolic-rate measurements in mammalian-cell culture by carbon and nitrogen balances. *Cytotechnology*, 29(3):167–176, 1999.
- Bozell, J. und Petersen, G. Technology development for the production of biobased products from biorefinery carbohydrates - the US Department of Energy's „Top 10“ revisited. *Green Chem.*, 12:539–54, 2010.
- Braatsch, S., Helmark, S., Kranz, H., Koebmann, B., und Jensen, P. R. *Escherichia coli* strains with promoter libraries constructed by Red/ET recombination pave the way for transcriptional fine-tuning. *Biotechniques*, 45(3):335–337, 2008.
- Broadhurst, D. I. und Kell, D. B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics*, 2(4):171–196, 2006.
- Burgard, A. P., Pharkya, P., und Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Bio-technol Bioeng*, 84(6):647–657, 2003. 077.

- Burgard, A. P., Nikolaev, E. V., Schilling, C. H., und Maranas, C. D. Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome Res*, 14(2):301–312, 2004.
- Carinhas, N., Bernal, V., Monteiro, F., Carrondo, M. J. T., Oliveira, R., and Alves, P. M. Improving baculovirus production at high cell density through manipulation of energy metabolism. *Metab Eng*, 12(1):39–52, 2010.
- Carinhas, N., Bernal, V., Teixeira, A. P., Carrondo, M. J., Alves, P. M., and Oliveira, R. Hybrid metabolic flux analysis: combining stoichiometric and statistical constraints to model the formation of complex recombinant products. *BMC Syst Biol*, 5:34, 2011.
- Carius, L., Hädicke, O., und Grammel, H. Stepwise reduction of the culture redox potential allows the analysis of microaerobic metabolism and photosynthetic membrane synthesis in *Rhodospirillum rubrum*. *Biotechnol Bioeng*, 110(2):573–585, 2013.
- Caspi, R., Altman, T., Dale, J. M., Dreher, K., Fulcher, C. A., Gilham, F., Kaipa, P., Karthikeyan, A. S., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L. A., Paley, S., Popescu, L., Pujar, A., Shearer, A. G., Zhang, P., und Karp, P. D. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res*, 38(Database issue):D473–D479, 2010.
- Chao, Y. P., Patnaik, R., Roof, W. D., Young, R. F., und Liao, J. C. Control of gluconeogenic growth by pps and pck in *Escherichia coli*. *J Bacteriol*, 175(21):6939–6944, Nov 1993.
- Chávez-Béjar, M. I., Lara, A. R., López, H., Hernández-Chávez, G., Martínez, A., Ramírez, O. T., Bolívar, F., und Gosset, G. Metabolic engineering of *Escherichia coli* for L-tyrosine production by expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from *Zymomonas mobilis*. *Appl Environ Microbiol*, 74(10):3284–3290, 2008. 176.
- Chemler, J. A., Fowler, Z. L., McHugh, K. P., und Koffas, M. A. G. Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering. *Metab Eng*, 12(2):96–104, Mar 2010.
- Chen, Z., Wilmanns, M., und Zeng, A.-P. Structural synthetic biotechnology: from molecular structure to predictable design for industrial strain development. *Trends Biotechnol*, 28(10):534–542, 2010.
- Cheng, K.-K., Zhao, X.-B., Zeng, J., und Zhang, J.-A. Biotechnological production of succinic acid: current state and perspectives. *Biofuels, Bioproducts and Biorefining*, 6(3):302–318, 2012.
- Choi, H. S., Lee, S. Y., Kim, T. Y., und Woo, H. M. In silico identification of gene amplification targets for improvement of lycopene production. *Appl Environ Microbiol*, 76(10):3097–3105, 2010.

- Cirino, P. C. und Arnold, F. H. Protein engineering of oxygenases for biocatalysis. *Curr Opin Chem Biol*, 6(2):130–135, 2002.
- Contador, C. A., Rizk, M. L., Asenjo, J. A., und Liao, J. C. Ensemble modeling for strain development of L-lysine-producing *Escherichia coli*. *Metab Eng*, 11(4-5):221–233, 2009. 281.
- Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M., Kim, K. H., Sleight, S. C., Maranas, C. D., und Sauro, H. M. Computational tools for metabolic engineering. *Metab Eng*, 14(3):270–280, 2012.
- Covert, M. W., Schilling, C. H., und Palsson, B. Ø. Regulation of gene expression in flux balance models of metabolism. *J Theor Biol*, 213(1):73–88, 2001.
- Covert, M. W., Knight, E. M., Reed, J. L., Herrgård, M. J., und Palsson, B. Ø. Integrating high-throughput and computational data elucidates bacterial networks. *Nature*, 429(6987):92–96, 2004.
- Covert, M. W., Xiao, N., Chen, T. J., und Karr, J. R. Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. *Bioinformatics*, 24(18): 2044–2050, 2008.
- Cramer, A., Raillard, S. A., Bermudez, E., und Stemmer, W. P. DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature*, 391(6664):288–291, 1998.
- Dantzig, G. *Linear programming and extension*. Princeton University Press, 1998.
- Datsenko, K. A. und Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A*, 97(12):6640–6645, 2000.
- Datta, R. und Henry, M. Lactic acid: recent advances in products, processes and technologies - a review. *J Chem Technol Biotechnol*, 81:1119 – 1129, 2006.
- Datta, R., Tsai, S., Bonsignore, P., Moon, S., und Frank, J. Technological and economic potential of poly(lactic acid) and lactic acid derivatives. *FEMS Microbiology Reviews*, 16:221–231, 1995.
- David, L., Marashi, S.-A., Larhlimi, A., Mieth, B., und Bockmayr, A. FFCA: a feasibility-based method for flux coupling analysis of metabolic networks. *BMC Bioinformatics*, 12:236, 2011.
- de Figueiredo, L. F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J. E., Schuster, S., und Planes, F. J. Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics*, 25(23):3158–3165, 2009.
- Demain, A. L. und Adrio, J. L. Strain improvement for production of pharmaceuticals and other microbial metabolites by fermentation. *Prog Drug Res*, 65:251, 253–251, 289, 2008.

- Demain, A. L. und Adrio, J. L. Essential role of genetics in the advancement of biotechnology. *Methods Mol Biol*, 898:1–40, 2012.
- Edwards, J. S. und Palsson, B. Ø. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. *BMC Bioinformatics*, 1:1, 2000a.
- Edwards, J. S. und Palsson, B. Ø. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A*, 97(10):5528–5533, 2000b. 141.
- Edwards, J. S., Ibarra, R. U., und Palsson, B. Ø. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat Biotechnol*, 19(2):125–130, 2001. 253.
- Elowitz, M. B., Levine, A. J., Siggia, E. D., und Swain, P. S. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–6, 2002. 054.
- Endy, D. Foundations for engineering biology. *Nature*, 438(7067):449–453, 2005.
- Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., und Palsson, B. Ø. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007. 113.
- Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., und Palsson, B. Ø. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 7(2):129–143, 2009.
- Fell, D. A. und Thomas, S. Physiological control of metabolic flux: the requirement for multisite modulation. *Biochem J*, 311 (Pt 1):35–39, 1995.
- Fong, S. S., Burgard, A. P., Herring, C. D., Knight, E. M., Blattner, F. R., Maranas, C. D., und Palsson, B. Ø. In silico design and adaptive evolution of Escherichia coli for production of lactic acid. *Biotechnol Bioeng*, 91(5):643–648, 2005. 142.
- Fong, S. S., Nanchen, A., Palsson, B. Ø., und Sauer, U. Latent pathway activation und increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes. *J Biol Chem*, 281(12):8024–8033, 2006. 148.
- Fukuda, K. und Prodon, A. Double Description Method Revisited. In *Combinatorics and Computer Science.*, volume 1120, pages 91–111. Springer, 1996.
- Gagneur, J. und Klamt, S. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, 5(175), 2004. 004.
- Gebauer, J., Schuster, S., de Figueiredo, L. F., und Kaleta, C. Detecting and investigating substrate cycles in a genome-scale human metabolic network. *FEBS J*, 279(17):3192–3202, 2012.
- Goldberg, I., Lonberg-Holm, K., Bagley, E. A., und Stieglitz, B. Improved conversion of fumarate to succinate by Escherichia coli strains amplified for fumarate reductase. *Appl Environ Microbiol*, 45(6):1838–1847, Jun 1983.

- Gong, J., Zheng, H., Wu, Z., Chen, T., und Zhao, X. Genome shuffling: Progress und applications for phenotype improvement. *Biotechnol Adv*, 27(6):996–1005, 2009. 282.
- Goudar, C. T., Biener, R., Konstantinov, K. B., und Piret, J. M. Error propagation from prime variables into specific rates and metabolic fluxes for mammalian cells in perfusion culture. *Biotechnol Prog*, 25(4):986–998, 2009.
- Gruchattka, E., Hädicke, O., Klamt, S., Schütz, V., und Kayser, O. In silico profiling of *Escherichia coli* and *Saccharomyces cerevisiae* as terpenoid factories. *Microb Cell Fact*, 12:84, 2013.
- Hädicke, O., Lohr, V., Genzel, Y., Reichl, U., und Klamt, S. Evaluating differences of metabolic performances: statistical methods and their application to animal cell cultivations. *Biotechnol Bioeng*, 110(10):2633–2642, 2013.
- Hädicke, O. und Klamt, S. CASOP: a computational approach for strain optimization aiming at high productivity. *J Biotechnol*, 147(2):88–101, 2010.
- Hädicke, O. und Klamt, S. Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab Eng*, 13(2):204–213, 2011.
- Hädicke, O., Grammel, H., und Klamt, S. Metabolic network modeling of redox balancing and biohydrogen production in purple nonsulfur bacteria. *BMC Syst Biol*, 5:150, 2011.
- Hammer, K., Mijakovic, I., und Jensen, P. R. Synthetic promoter libraries—tuning of gene expression. *Trends Biotechnol*, 24(2):53–55, 2006.
- Haus, U.-U., Klamt, S., und Stephen, T. Computing knock-out strategies in metabolic networks. *J Comput Biol*, 15(3):259–268, 2008. 149.
- Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Lindsay, B., und Stevens, R. L. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*, 28(9):977–982, 2010.
- Hermann, B. G., Blok, K., und Patel, M. K. Producing bio-based bulk chemicals using industrial biotechnology saves energy and combats climate change. *Environ Sci Technol*, 41(22):7915–7921, Nov 2007.
- Hilfinger, A. und Paulsson, J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci U S A*, 108(29):12167–12172, 2011.
- Hofvendahl und Hahn-Hägerdal. Factors affecting the fermentative lactic acid production from renewable resources(1). *Enzyme Microb Technol*, 26(2-4):87–107, 2000.
- Holm, A. K., Blank, L. M., Oldiges, M., Schmid, A., Solem, C., Jensen, P. R., und Vemuri, G. N. Metabolic and transcriptional response to cofactor perturbations in *Escherichia coli*. *J Biol Chem*, 285(23):17498–17506, Jun 2010.
- Hoppe, A., Hoffmann, S., und Holzhütter, H.-G. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC Syst Biol*, 1:23, 2007. 070.

- Hua, Q., Joyce, A. R., Fong, S. S., und Palsson, B. Ø. Metabolic analysis of adaptive evolution for in silico-designed lactate-producing strains. *Biotechnol Bioeng*, 95(5): 992–1002, 2006. 098.
- Ibarra, R. U., Edwards, J. S., und Palsson, B. Ø. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature*, 420(6912):186–189, 2002. 167.
- Ip, K., Colijn, C., und Lun, D. S. Analysis of complex metabolic behavior through pathway decomposition. *BMC Syst Biol*, 5:91, 2011.
- Jamshidi, N. und Palsson, B. Ø. Formulating genome-scale kinetic models in the post-genome era. *Mol Syst Biol*, 4:171, 2008. 150.
- Jantama, K., Haupt, M. J., Svoronos, S. A., Zhang, X., Moore, J. C., Shanmugam, K. T., und Ingram, L. O. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of Escherichia coli C that produce succinate and malate. *Biotechnol Bioeng*, 99(5):1140–1153, 2008a. 100.
- Jantama, K., Zhang, X., Moore, J. C., Shanmugam, K. T., Svoronos, S. A., und Ingram, L. O. Eliminating side products and increasing succinate yields in engineered strains of Escherichia coli C. *Biotechnol Bioeng*, 101(5):881–893, 2008b. 182.
- Jensen, P. R. und Hammer, K. Artificial promoters for metabolic optimization. *Biotechnol Bioeng*, 58(2-3):191–195, 1998a.
- Jensen, P. R. und Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. *Appl Environ Microbiol*, 64(1):82–87, 1998b.
- Jin, Y.-S. und Stephanopoulos, G. Multi-dimensional gene target search for improving lycopene biosynthesis in Escherichia coli. *Metab Eng*, 9(4):337–347, 2007. 099.
- Johnson, I. S. Human insulin from recombinant DNA technology. *Science*, 219(4585): 632–637, 1983.
- Jung, Y. K., Kim, T. Y., Park, S. J., und Lee, S. Y. Metabolic engineering of Escherichia coli for the production of polylactic acid and its copolymers. *Biotechnol Bioeng*, 105(1):161–171, 2010.
- Jungreuthmayer, C. und Zanghellini, J. Designing optimal cell factories: integer programming couples elementary mode analysis with regulation. *BMC Syst Biol*, 6:103, 2012.
- Jungreuthmayer, C., Nair, G., Klamt, S., und Zanghellini, J. Comparison and improvement of algorithms for computing minimal cut sets. *BMC Bioinformatics*, 14:318, 2013.
- Kaleta, C., de Figueiredo, L. F., Behre, J., und Schuster, S. EFMEvolver: Computing elementary flux modes in genome-scale metabolic networks. *Lecture Notes in Informatics P-157*, pages 179–189, 2009.

- Kanehisa, M. A database for post-genome analysis. *Trends Genet*, 13(9):375–376, 1997.
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., und Tanabe, M. Data, information, knowledge und principle: back to metabolism in KEGG. *Nucleic Acids Res*, 42(Database issue):D199–D205, 2014.
- Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M., Latendresse, M., Muñoz-Rascado, L., Ong, Q., Paley, S., Schröder, I., Shearer, A. G., Subhraveti, P., Travers, M., Weerasinghe, D., Weiss, V., Collado-Vides, J., Gunsalus, R. P., Paulsen, I., und Karp, P. D. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res*, 41(Database issue):D605–D612, 2013.
- Khannapho, C., Zhao, H., Bonde, B. K., Kierzek, A. M., Avignone-Rossa, C. A., und Bushell, M. E. Selection of objective function in genome scale flux balance analysis for process feed development in antibiotic production. *Metab Eng*, 10(5):227–233, 2008.
- Kim, H.-J., Kwon, Y. D., Lee, S. Y., und Kim, P. An engineered Escherichia coli having a high intracellular level of ATP and enhanced recombinant protein production. *Appl Microbiol Biotechnol*, 94(4):1079–1086, May 2012a.
- Kim, J. I., Song, H.-S., Sunkara, S. R., Lali, A., und Ramkrishna, D. Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat. *Biotechnol Prog*, 28(5):1160–1166, 2012b.
- Kim, J. und Reed, J. L. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst Biol*, 4:53, 2010.
- Kim, J., Reed, J. L., und Maravelias, C. T. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. *PLoS One*, 6(9):e24162, 2011.
- Kim, P., Laivenieks, M., Vieille, C., und Zeikus, J. G. Effect of overexpression of Actinobacillus succinogenes phosphoenolpyruvate carboxykinase on succinate production in Escherichia coli. *Appl Environ Microbiol*, 70(2):1238–1241, Feb 2004.
- Klamt, S., Gagneur, J., und von Kamp, A. Algorithmic approaches for computing elementary modes in large biochemical reaction networks. *IEE Proc Syst Biol*, 152(4): 249–55, 2005. 060.
- Klamt, S. Generalized concept of minimal cut sets in biochemical networks. *BioSystems*, 83(1):233–247, 2006. 001.
- Klamt, S. und Gilles, E. D. Minimal cut sets in biochemical reaction networks. *Bioinformatics*, 20(2):226–234, 2004. 003.
- Klamt, S. und Stelling, J. Combinatorial complexity of pathway analysis in metabolic networks. *Mol Biol Rep*, 29(1-2):233–236, 2002.
- Klamt, S. und Stelling, J. Two approaches for metabolic pathway analysis? *Trends Biotechnol*, 21(2):64–9, 2003. 073.

- Klamt, S. und von Kamp, A. An application programming interface for CellNetAnalyzer. *Biosystems*, 105(2):162–168, 2011.
- Klamt, S., Schuster, S., und Gilles, E. D. Calculability Analysis in Underdetermined Metabolic Networks Illustrated by a Model of the Central Metabolism in Purple Nonsulfur Bacteria. *Biotechnol Bioeng.*, 77(7):734–751, 2002. 002.
- Klamt, S., Saez-Rodriguez, J., Lindquist, J. A., Simeoni, L., und Gilles, E. D. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics*, 7(56), 2006. 005.
- Klamt, S., Saez-Rodriguez, J., und Gilles, E. D. Structural und functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst Biol*, 1:2, 2007. 165.
- Klamt, S., Hädicke, O., und von Kamp, A. Stoichiometric and constraint-based analysis of biochemical reaction networks. In *Large-Scale Networks in Engineering and Life Sciences*. Springer, 2014.
- Koebmann, B. J., Westerhoff, H. V., Snoep, J. L., Nilsson, D., und Jensen, P. R. The glycolytic flux in *Escherichia coli* is controlled by the demand for ATP. *J Bacteriol*, 184(14):3909–3916, Jul 2002.
- Kondo, A., Ishii, J., Hara, K. Y., Hasunuma, T., und Matsuda, F. Development of microbial cell factories for bio-refinery through synthetic bioengineering. *J Biotechnol*, 163(2): 204–216, 2013.
- Korkegian, A., Black, M. E., Baker, D., und Stoddard, B. L. Computational thermostabilization of an enzyme. *Science*, 308(5723):857–860, 2005.
- Lee, J. W., Kim, T. Y., Jang, Y.-S., Choi, S., und Lee, S. Y. Systems metabolic engineering for chemicals and materials. *Trends Biotechnol*, 29(8):370–378, 2011.
- Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., und Lee, S. Y. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat Chem Biol*, 8(6):536–546, 2012a.
- Lee, K. H., Park, J. H., Kim, T. Y., Kim, H. U., und Lee, S. Y. Systems metabolic engineering of *Escherichia coli* for L-threonine production. *Mol Syst Biol*, 3:149, 2007. 178.
- Lee, S. Y., Lee, D.-Y., und Kim, T. Y. Systems biotechnology for strain improvement. *Trends Biotechnol*, 23(7):349–358, 2005.
- Lee, S. Y., Kim, H. U., Park, J. H., Park, J. M., und Kim, T. Y. Metabolic engineering of microorganisms: general strategies and drug production. *Drug Discov Today*, 14(1-2): 78–88, 2009. 181.
- Lee, S. Y., Mattanovich, D., und Villaverde, A. Systems metabolic engineering, industrial biotechnology and microbial cell factories. *Microb Cell Fact*, 11:156, 2012b.

- Leiser, J. und Blum, J. J. On the analysis of substrate cycles in large metabolic systems. *Cell Biophys*, 11:123–138, 1987.
- Leisola, M. und Turunen, O. Protein engineering: opportunities and challenges. *Appl Microbiol Biotechnol*, 75(6):1225–1232, 2007.
- Lewis, N. E., Nagarajan, H., und Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*, 10(4):291–305, 2012.
- Lin, H., Bennett, G. N., und San, K.-Y. Genetic reconstruction of the aerobic central metabolism in *Escherichia coli* for the absolute aerobic production of succinate. *Biotechnol Bioeng*, 89(2):148–156, 2005a. 136.
- Lin, H., San, K.-Y., und Bennett, G. N. Effect of *Sorghum vulgare* phosphoenolpyruvate carboxylase and *Lactococcus lactis* pyruvate carboxylase coexpression on succinate production in mutant strains of *Escherichia coli*. *Appl Microbiol Biotechnol*, 67(4):515–523, 2005b. 137.
- Llaneras, F. und Picó, J. A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. *BMC Bioinformatics*, 8:421, 2007. 090.
- Lohr, V., Hädicke, O., Genzel, Y., Jordan, I., Büntemeyer, H., Klamt, S., und Reichl, U. The avian cell line AGE1.CR.pIX characterized by metabolic flux analysis. *BMC Biotechnol*, 14:72, 2014.
- Lu, S., Eiteman, M. A., und Altman, E. Effect of CO₂ on succinate production in dual-phase *Escherichia coli* fermentations. *J Biotechnol*, 143(3):213–223, Sep 2009.
- Lun, D. S., Rockwell, G., Guido, N. J., Baym, M., Kelner, J. A., Berger, B., Galagan, J. E., und Church, G. M. Large-scale identification of genetic design strategies using local search. *Mol Syst Biol*, 5:296, 2009.
- Lynch, M. D., Warnecke, T., und Gill, R. T. SCALES: multiscale analysis of library enrichment. *Nat Methods*, 4(1):87–93, 2007.
- Machado, D., Costa, R. S., Ferreira, E. C., Rocha, I., und Tidor, B. Exploring the gap between dynamic and constraint-based models of metabolism. *Metab Eng*, 14(2):112–119, 2012a.
- Machado, D., Soons, Z., Patil, K. R., Ferreira, E. C., und Rocha, I. Random sampling of elementary flux modes in large-scale metabolic networks. *Bioinformatics*, 28(18):i515–i521, 2012b.
- Maertens, J. und Vanrolleghem, P. A. Modeling with a view to target identification in metabolic engineering: a critical evaluation of the available tools. *Biotechnol Prog*, 26(2):313–331, 2010.
- Mahadevan, R. und Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*, 5(4):264–276, 2003.

- Mahadevan, R., Edwards, J. S., und Doyle, F. J. 3rd. Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys J*, 83(3):1331–1340, 2002.
- Makino, T., Skretas, G., und Georgiou, G. Strain engineering for improved expression of recombinant proteins in bacteria. *Microb Cell Fact*, 10:32, 2011.
- Marashi, S.-A., David, L., und Bockmayr, A. Analysis of metabolic subnetworks by flux cone projection. *Algorithms Mol Biol*, 7(1):17, 2012.
- Martinez-Morales, F., Borges, A. C., Martinez, A., Shanmugam, K. T., und Ingram, L. O. Chromosomal integration of heterologous DNA in *Escherichia coli* with precise removal of markers and replicons used during construction. *J Bacteriol*, 181(22):7143–7148, 1999.
- Maskow, T. und von Stockar, U. How reliable are thermodynamic feasibility statements of biochemical pathways? *Biotechnol Bioeng*, 92(2):223–230, 2005.
- Mavrovouniotis, M. L., Stephanopoulos, G., und Stephanopoulos, G. Computer-aided synthesis of biochemical pathways. *Biotechnol Bioeng*, 36(11):1119–1132, 1990.
- Melzer, G., Esfandabadi, M. E., Franco-Lara, E., und Wittmann, C. Flux Design: In silico design of cell factories based on correlation of pathway fluxes to desired properties. *BMC Syst Biol*, 3:120, 2009.
- Millard, C. S., Chao, Y. P., Liao, J. C., und Donnelly, M. I. Enhanced production of succinic acid by overexpression of phosphoenolpyruvate carboxylase in *Escherichia coli*. *Appl Environ Microbiol*, 62(5):1808–1810, May 1996.
- Na, D., Kim, T. Y., und Lee, S. Y. Construction and optimization of synthetic pathways in metabolic engineering. *Curr Opin Microbiol*, 13(3):363–370, 2010.
- Niederberger, P., Prasad, R., Miozzari, G., und Kacser, H. A strategy for increasing an in vivo flux by genetic manipulations. The tryptophan system of yeast. *Biochem J*, 287 (Pt 2):473–479, 1992.
- Nielsen, J. Principles of optimal metabolic network operation. *Mol Syst Biol*, 3(126), 2007. 017.
- Nusser, M., Hüsing, B., und Wydra, S. Potenzialanalyse der industriellen, weißen Biotechnologie. *FraunhoferInstitut für System-und Innovationsforschung: Karlsruhe, Germany*, 2007.
- Orth, J. D., Conrad, T. M., Na, J., Lerman, J. A., Nam, H., Feist, A. M., and Palsson, B. Ø. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism–2011. *Mol Syst Biol*, 7:535, 2011.
- Pande, S., Merker, H., Bohl, K., Reichelt, M., Schuster, S., de Figueiredo, L. F., Kaleta, C., und Kost, C. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J*, 8(5):953–962, May 2014.

- Papagianni, M. Recent advances in engineering the central carbon metabolism of industrially important bacteria. *Microb Cell Fact*, 11:50, 2012.
- Papin, J. A., Stelling, J., Price, N. D., Klamt, S., Schuster, S., und Palsson, B. Ø. Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8):400–5, 2004. 074.
- Park, J. H., Lee, S. Y., Kim, T. Y., und Kim, H. U. Application of systems biology for bioprocess development. *Trends Biotechnol*, 26(8):404–412, 2008.
- Park, J. M., Park, H. M., Kim, W. J., Kim, H. U., Kim, T. Y., und Lee, S. Y. Flux variability scanning based on enforced objective flux for identifying gene amplification targets. *BMC Syst Biol*, 6:106, 2012.
- Patil, K. R., Rocha, I., Förster, J., und Nielsen, J. Evolutionary programming as a platform for in silico metabolic engineering. *BMC Bioinformatics*, 6:308, 2005. 084.
- Patnaik, R., Roof, W. D., Young, R. F., und Liao, J. C. Stimulation of glucose catabolism in *Escherichia coli* by a potential futile cycle. *J Bacteriol*, 174(23):7527–7532, Dec 1992.
- Patnaik, R. Engineering complex phenotypes in industrial strains. *Biotechnol Prog*, 24(1):38–47, 2008.
- Patnaik, R., Louie, S., Gavrilovic, V., Perry, K., Stemmer, W. P. C., Ryan, C. M., und del Cardayré, S. Genome shuffling of *Lactobacillus* for improved acid tolerance. *Nat Biotechnol*, 20(7):707–712, 2002.
- Petrounia, I. P. und Arnold, F. H. Designed evolution of enzymatic properties. *Curr Opin Biotechnol*, 11(4):325–330, 2000.
- Pfeiffer, T., Sánchez-Valdenebro, I., no, J. C. N., Montero, F., und Schuster, S. META-TOOL: for studying metabolic networks. *Bioinformatics*, 15(3):251–257, 1999. 168.
- Pfleger, B. F., Pitera, D. J., Smolke, C. D., und Keasling, J. D. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol*, 24(8):1027–1032, 2006.
- Pharkya, P. und Maranas, C. D. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab Eng*, 8(1):1–13, 2006. 081.
- Pharkya, P., Burgard, A. P., und Maranas, C. D. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res*, 14(11):2367–2376, 2004. 078.
- Pósfai, G., Koob, M. D., Kirkpatrick, H. A., und Blattner, F. R. Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the *Escherichia coli* O157:H7 genome. *J Bacteriol*, 179(13):4426–4428, 1997.

- Price, N. D., Papin, J. A., Schilling, C. H., und Palsson, B. Ø. Genome-scale microbial in silico models: the constraints-based approach. *Trends Biotechnol*, 21(4):162–9, 2003. 065.
- Price, N. D., Reed, J. L., und Palsson, B. Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*, 2(11):886–97, 2004. 064.
- Ranganathan, S., Suthers, P. F., und Maranas, C. D. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput Biol*, 6(4):e1000744, 2010.
- Rausenberger, J. und Kollmann, M. Quantifying origins of cell-to-cell variations in gene expression. *Biophys J*, 95(10):4523–4528, 2008.
- Reed, J. L. Shrinking the metabolic solution space using experimental datasets. *PLoS Comput Biol*, 8(8):e1002662, 2012.
- Reed, J. L. und Palsson, B. Ø. Genome-scale in silico models of E. coli have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res*, 14(9):1797–1805, 2004.
- Reed, J. L., Vo, T. D., Schilling, C. H., und Palsson, B. O. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol*, 4(9):R54, 2003.
- Reed, J. L., Senger, R. S., Antoniewicz, M. R., und Young, J. D. Computational approaches in metabolic engineering. *J Biomed Biotechnol*, 2010:207414, 2010.
- Rizk, M. L. und Liao, J. C. Ensemble modeling for aromatic production in Escherichia coli. *PLoS One*, 4(9):e6903, 2009.
- Rockafellar, R. *Conex Analysis*. University Press, 1970.
- San, K.-Y., Bennett, G. N., Berríos-Rivera, S. J., Vadali, R. V., Yang, Y.-T., Horton, E., Rudolph, F. B., Sariyar, B., und Blackwood, K. Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in Escherichia coli. *Metab Eng*, 4(2):182–192, Apr 2002.
- Sánchez, A. M., Bennett, G. N., und San, K.-Y. Novel pathway engineering design of the anaerobic central metabolic pathway in Escherichia coli to increase succinate yield and productivity. *Metab Eng*, 7(3):229–239, 2005. 105.
- Sánchez, A. M., Bennett, G. N., und San, K.-Y. Batch culture characterization and metabolic flux analysis of succinate-producing Escherichia coli strains. *Metab Eng*, 8(3): 209–226, 2006. 108.
- Sauer, M., Porro, D., Mattanovich, D., und Branduardi, P. Microbial production of organic acids: expanding the markets. *Trends Biotechnol*, 26(2):100–108, 2008.

- Sauer, M., Porro, D., Mattanovich, D., und Branduardi, P. 16 years research on lactic acid production with yeast - ready for the market? *Biotechnol Genet Eng Rev*, 27:229–256, 2010.
- Sauer, U. Evolutionary engineering of industrially important microbial phenotypes. *Adv Biochem Eng Biotechnol*, 73:129–169, 2001.
- Schellenberger, J., Lewis, N. E., und Palsson, B. Ø. Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophys J*, 100(3):544–553, 2011.
- Schilling, C. H., Letscher, D., und Palsson, B. O. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol*, 203(3):229–248, 2000.
- Schuetz, R., Kuepfer, L., und Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Mol Syst Biol*, 3(119), 2007. 010.
- Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M., und Sauer, U. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, 2012.
- Schuster, S., Fell, D. A., und Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol*, 18(3):326–332, 2000.
- Schuster, S. und Hilgetag, C. On Elementary Flux Modes in Biochemical Reaction Systems at Steady State. *Journal of Biological Systems*, 2:165–182, 1994.
- Schuster, S., Pfeiffer, T., und Fell, D. A. Is maximization of molar yield in metabolic networks favoured by evolution? *J Theor Biol*, 252(3):497–504, 2008. 166.
- Segrè, D., Vitkup, D., und Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*, 99(23):15112–15117, 2002.
- Shlomi, T., Berkman, O., und Ruppin, E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A*, 102(21):7695–7700, 2005. 278.
- Shlomi, T., Eisenberg, Y., Sharan, R., und Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*, 3:101, 2007. 072.
- Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø., und Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol*, 26(9):1003–1010, 2008.
- Singh, A., Cher Soh, K., Hatzimanikatis, V., und Gill, R. T. Manipulating redox and ATP balancing for improved production of succinate in E. coli. *Metab Eng*, 13(1):76–81, Jan 2011.
- Soetaert, W. und Vandamme, E. J. *Industrial Biotechnology. Sustainable Growth and Economic Success*. Wiley-VCH, Weinheim, 2010.

- Solem, C. und Jensen, P. R. Modulation of gene expression made easy. *Appl Environ Microbiol*, 68(5):2397–2403, 2002.
- Song, H. und Lee, S. Y. Production of succinic acid by bacterial fermentation. *Enzyme and microbial technology*, 39(3):352–361, 2006.
- Song, H.-S. und Ramkrishna, D. Prediction of dynamic behavior of mutant strains from limited wild-type data. *Metab Eng*, 14(2):69–80, 2012.
- Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., und Gilles, E. D. Metabolic network structure determines key aspects of functionality and regulation. *Nature*, 420(6912): 190–193, 2002. 037.
- Stephanopoulos, G., Aristidou, A., und Nielsen, J. *Metabolic Engineering*. Academic Press, San Diego, 1998.
- Stephanopoulos, G. und Stafford, D. E. Metabolic Engineering: a new frontier of chemical reaction engineering. *Chemical Engineering Science*, 57:2595–2602, 2002.
- Suthers, P. F., Zomorodi, A., und Maranas, C. D. Genome-scale gene/reaction essentiality and synthetic lethality analysis. *Mol Syst Biol*, 5:301, 2009.
- Tan, Y. und Liao, J. C. Metabolic ensemble modeling for strain engineers. *Biotechnol J*, 7(3):343–353, 2012.
- Tepper, N. und Shlomi, T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*, 26(4):536–543, 2010. 292.
- Terzer, M. und Stelling, J. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, 24(19):2229–2235, 2008.
- Teusink, B., Wiersma, A., Jacobs, L., Notebaart, R. A., und Smid, E. J. Understanding the adaptive growth strategy of *Lactobacillus plantarum* by in silico optimisation. *PLoS Comput Biol*, 5(6):e1000410, Jun 2009.
- Thakker, C., Martínez, I., San, K.-Y., und Bennett, G. N. Succinate production in *Escherichia coli*. *Biotechnol J*, 7(2):213–224, Feb 2012.
- Thiele, I. und Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*, 5(1):93–121, 2010.
- Tomar, N. und De, R. K. Comparing methods for metabolic network analysis and an application to metabolic engineering. *Gene*, 521(1):1–14, 2013.
- Tran, L. M., Rizk, M. L., und Liao, J. C. Ensemble modeling of metabolic networks. *Biophys J*, 95(12):5606–5617, 2008.
- Trinh, C. T. und Srienc, F. Metabolic engineering of *Escherichia coli* for efficient conversion of glycerol to ethanol. *Appl Environ Microbiol*, 75(21):6696–6705, 2009. 288.

- Trinh, C. T., Carlson, R., Wlaschin, A., und Srienc, F. Design, construction und performance of the most efficient biomass producing *E. coli* bacterium. *Metab Eng*, 8(6): 628–638, 2006. 101.
- Trinh, C. T., Unrean, P., und Srienc, F. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl Environ Microbiol*, 74(12): 3634–3643, 2008. 145.
- Trinh, C. T., Wlaschin, A., und Srienc, F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol*, 81(5):813–826, 2009. 164.
- Unrean, P., Trinh, C. T., und Srienc, F. Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. *Metab Eng*, 12(2):112–122, 2010. 290.
- Urbanczik, R. und Wagner, C. An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics*, 21(7):1203–1210, 2005.
- Vadali, R. V., Fu, Y., Bennett, G. N., und San, K.-Y. Enhanced lycopene productivity by manipulation of carbon flow to isopentenyl diphosphate in *Escherichia coli*. *Biotechnol Prog*, 21(5):1558–1561, 2005. 156.
- van Berlo, R. J. P., de Ridder, D., Daran, J.-M., Daran-Lapujade, P. A. S., Teusink, B., und Reinders, M. J. T. Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM Trans Comput Biol Bioinform*, 8(1):206–216, 2011.
- van der Heijden, R. T., Romein, B., Heijnen, J. J., Hellinga, C., und Luyben, K. C. Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross errors. *Biotechnol Bioeng*, 43(1):11–20, 1994.
- Varma, A. und Palsson, B. Ø. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl Environ Microbiol*, 60(10):3724–3731, 1994. 116.
- Varma, A., Boesch, B. W., und Palsson, B. O. Biochemical production capabilities of *Escherichia coli*. *Biotechnol Bioeng*, 42(1):59–73, 1993.
- Vemuri, G. N., Eiteman, M. A., und Altman, E. Succinate production in dual-phase *Escherichia coli* fermentations depends on the time of transition from aerobic to anaerobic conditions. *J Ind Microbiol Biotechnol*, 28(6):325–332, Jun 2002a.
- Vemuri, G. N., Eiteman, M. A., und Altman, E. Effects of growth mode und pyruvate carboxylase on succinic acid production by metabolically engineered strains of *Escherichia coli*. *Appl Environ Microbiol*, 68(4):1715–1727, Apr 2002b.
- Vink, E. T., Glassner, D., R.J., K., Wooley, R., und OConnor, R. The eco-profiles for current and near-future NatureWorks polylactide (PLA) production. *Industrial Biotechnology*, 3(1):58–81, 2007.

- Vojinović, V. und von Stockar, U. Influence of uncertainties in pH, pMg, activity coefficients, metabolite concentrations, and other factors on the analysis of the thermodynamic feasibility of metabolic pathways. *Biotechnol Bioeng*, 103(4):780–795, 2009.
- von Kamp, A. und Klamt, S. Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Comput Biol*, 10(1):e1003378, 2014.
- von Kamp, A. und Schuster, S. Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, 22(15):1930–1931, 2006.
- von Stockar, U. *Thermodynamics in Biochemical Engineering*. EPFL Press, 2013. ISBN 1466582162.
- von Stockar, U., Maskow, T., Liu, J., Marison, I. W., und Patiño, R. Thermodynamics of microbial growth and metabolism: an analysis of the current situation. *J Biotechnol*, 121(4):517–533, 2006.
- Wang, C.-w., Oh, M.-K., und Liao, J. C. Directed Evolution of Metabolically Engineered *Escherichia coli* for Carotenoid Production. *Biotechnology progress*, 16(6):922–926, 2000.
- Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–898, 2009. 262.
- Wang, L., Birol, I., und Hatzimanikatis, V. Metabolic control analysis under uncertainty: framework development and case studies. *Biophys J*, 87(6):3750–3763, 2004.
- Wang, X., Gong, C., und Tsao, G. T. Bioconversion of fumaric acid to succinic acid by recombinant *E. coli*. *Applied biochemistry and biotechnology*, 70(1):919–928, 1998.
- Warner, J. R., Patnaik, R., und Gill, R. T. Genomics enabled approaches in strain engineering. *Curr Opin Microbiol*, 12(3):223–230, 2009.
- Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A., und Gill, R. T. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol*, 28(8):856–862, 2010.
- Wendisch, V. F., Bott, M., und Eikmanns, B. J. Metabolic engineering of *Escherichia coli* and *Corynebacterium glutamicum* for biotechnological production of organic acids and amino acids. *Curr Opin Microbiol*, 9(3):268–274, 2006. 103.
- Werner, S., Diekert, G., und Schuster, S. Revisiting the thermodynamic theory of optimal ATP stoichiometries by analysis of various ATP-producing metabolic pathways. *J Mol Evol*, 71(5-6):346–355, 2010.
- Weusthuis, R. A., Lamot, I., van der Oost, J., und Sanders, J. P. M. Microbial production of bulk chemicals: development of anaerobic processes. *Trends Biotechnol*, 29(4):153–158, Apr 2011.

- Wiback, S. J., Mahadevan, R., und Palsson, B. Ø. Reconstructing metabolic flux vectors from extreme pathways: defining the alpha-spectrum. *J Theor Biol*, 224(3):313–324, 2003.
- Wiback, S. J., Famili, I., Greenberg, H. J., und Palsson, B. Ø. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J Theor Biol*, 228(4):437–47, 2004. 040.
- Wiechert, W. ¹³C metabolic flux analysis. *Metab Eng*, 3(3):195–206, 2001. 197.
- Wiechert, W. und de Graaf, A. A. In vivo stationary flux analysis by ¹³C labeling experiments. *Adv Biochem Eng Biotechnol*, 54:109–154, 1996.
- Wiechert, W., Möllney, M., Petersen, S., und de Graaf, A. A. A universal framework for ¹³C metabolic flux analysis. *Metab Eng*, 3(3):265–283, 2001. 169.
- Wiechert, W., Siefke, C., de Graaf, A. A., und Marx, A. Bidirectional Reaction Steps in Metabolic Networks Part II: Flux Estimation and Statistical Analysis. *Biotechnol Bioeng*, 55:118–135, 1997. 021.
- Wittmann, C. und Lee, S. Y. *Systems Metabolic Engineering*. Springer, 2012.
- Yang, J., Wang, Z., Zhu, N., Wang, B., Chen, T., und Zhao, X. Metabolic engineering of *Escherichia coli* and in silico comparing of carboxylation pathways for high succinate productivity under aerobic conditions. *Microbiol Res*, 2013.
- Yang, L., Cluett, W. R., und Mahadevan, R. EMILiO: a fast algorithm for genome-scale strain design. *Metab Eng*, 13(3):272–281, 2011.
- Yoon, S.-H., Lee, Y.-M., Kim, J.-E., Lee, S.-H., Lee, J.-H., Kim, J.-Y., Jung, K.-H., Shin, Y.-C., Keasling, J. D., und Kim, S.-W. Enhanced lycopene production in *Escherichia coli* engineered to synthesize isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate. *Biotechnol Bioeng*, 94(6):1025–1032, 2006. 152.
- Young, J. D., Henne, K. L., Morgan, J. A., Konopka, A. E., und Ramkrishna, D. Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. *Biotechnol Bioeng*, 100(3):542–559, 2008. 265.
- Yu, C., Cao, Y., Zou, H., und Xian, M. Metabolic engineering of *Escherichia coli* for biotechnological production of high-value organic acids and alcohols. *Appl Microbiol Biotechnol*, 89(3):573–583, 2011.
- Zhang, X., Jantama, K., Moore, J. C., Shanmugam, K. T., und Ingram, L. O. Production of L-alanine by metabolically engineered *Escherichia coli*. *Appl Microbiol Biotechnol*, 77(2):355–366, 2007.
- Zhang, Y.-X., Perry, K., Vinci, V. A., Powell, K., Stemmer, W. P. C., und del Cardayré, S. B. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature*, 415(6872):644–646, 2002.

- Zhao, H., Giver, L., Shao, Z., Affholter, J. A., und Arnold, F. H. Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nat Biotechnol*, 16(3): 258–261, 1998.
- Zhou, S., Yomano, L. P., Shanmugam, K. T., und Ingram, L. O. Fermentation of 10Escherichia coli B. *Biotechnol Lett*, 27(23-24):1891–1896, 2005. 143.
- Zhou, S., Causey, T., Hasona, A., Shanmugam, K., und Ingram, L. Production of optically pure D-lactic acid in mineral salts medium by metabolically engineered Escherichia coli W3110. *Applied and environmental microbiology*, 69(1):399–407, 2003.
- Zhou, S., Ingram, L. O., Shanmugam, K. T., Yomano, L., Grabar, T. B., and Moore, J. C. Materials and methods for efficient lactic acid production, April 23 2013. URL <http://www.google.com/patents/US8426191>. US Patent 8,426,191.
- Zhuang, K., Yang, L., Cluett, W. R., und Mahadevan, R. Dynamic strain scanning optimization: an efficient strain design strategy for balanced yield, titer, and productivity. DySScO strategy for strain design. *BMC Biotechnol*, 13:8, 2013.
- Zomorodi, A. R., Suthers, P. F., Ranganathan, S., und Maranas, C. D. Mathematical optimization applications in metabolic networks. *Metab Eng*, 14(6):672–686, 2012.
- Zomorodi, A. R., Lafontaine Rivera, J. G., Liao, J. C., und Maranas, C. D. Optimization-driven identification of genetic perturbations accelerates the convergence of model parameters in ensemble modeling of metabolic networks. *Biotechnol J*, 8(9):1090–1104, 2013.