

Entwicklung neuer Methoden zur Integration thermodynamischer und enzymatischer Nebenbedingungen in der constraint-basierten metabolischen Modellierung

Dissertation

zur Erlangung des akademischen Grades

Doktoringenieur (Dr.-Ing.)

von M.Sc. Pavlos Stephanos Bekiaris
geb. am 4. Dezember 1994 in Herford (Nordrhein-Westfalen)

genehmigt durch die Fakultät für Verfahrens- und Systemtechnik der Otto-von-Guericke-
Universität Magdeburg

Gutachter:

Dr.-Ing. Steffen Klamt

Prof. Dr.-Ing. Oliver Peters

Univ.-Prof. DI Dr. Jürgen Zanghellini

Promotionskolloquium am 22. Mai 2024

Zusammenfassung

Mithilfe der constraint-basierten Modellierung ist es möglich, zentrale Aspekte des Stoffwechsels von Organismen mathematisch abzubilden, zu analysieren und vorherzusagen. In den letzten 4 Jahrzehnten hat sich dabei eine Vielzahl an Methoden der constraint-basierten Modellierung entwickelt, die ihrerseits auf eine Vielzahl biochemischer und physikalischer Daten zurückgreifen können.

In dieser Arbeit werden drei neue Methoden der constraint-basierten Modellierung vorgestellt: Die erste neue Methode ist sMOMENT, mit der enzymatische Nebenbedingungen in constraint-basierte Modelle integriert werden können, wobei nur eine minimale Erweiterung des Modells erforderlich ist. Dadurch kann die Größe metabolischer Modelle mit enzymatischen Nebenbedingungen im Vergleich zu entsprechenden bisherigen Methoden deutlich reduziert werden. Hiervon profitieren insbesondere rechenintensive Analysen und Methoden. Als Referenzimplementierung von sMOMENT wurde zudem das Programmpaket AutoPACMEN entwickelt, mit dem es erstmals möglich wurde, gleichzeitig relevante enzymatische Daten zu sammeln, und daraus ein Modell mit enzymatischen Nebenbedingungen zu erstellen und zu kalibrieren. AutoPACMEN und sMOMENT wurden exemplarisch an einem genomskaligen Modell von *Escherichia coli* angewendet. Das so erstellte Modell kann unter anderem sowohl unter aeroben, als auch anaeroben Bedingungen den Überlaufstoffwechsel von *E. coli* erfolgreich vorhersagen. Die zweite neue Methode ist TCOSA. Sie verwendet das Konzept der netzwerkweiten thermodynamischen Triebkräfte um die Relevanz redundanter Cofaktorpaare zu analysieren. Durch die Anwendung von TCOSA auf ein genomskaliges Modell von *E. coli* und dessen NAD(P)(H)-Reaktionsspezifitäten konnte gezeigt werden, dass die Evolution dieser Spezifitäten offensichtlich stark durch thermodynamische Einschränkungen beeinflusst wurde. Darüber hinaus konnten mit diesem Modell *in vivo* auftretende qualitative Trends der Redoxzustände von NAD(H) sowie NADP(H) vorhergesagt werden. Weiterhin konnte gezeigt werden, dass die Anwesenheit von mehr als zwei Cofaktorpaaren nur dann thermodynamisch sinnvoll erscheint, wenn das Standard-Redoxpotential des dritten Cofaktorpaares im Vergleich zu NAD(P)(H) erniedrigt ist. Die dritte neue Methode ist ASTHERISC. Mit ihrer Hilfe können thermodynamisch vorteilhafte Produktsynthesewege in Gemeinschaften mehrerer Stämme einer Spezies im Vergleich zu Einzelstämmen vorhergesagt werden. ASTHERISC basiert ebenfalls auf netzwerkweiten thermodynamischen Triebkräften und wurde exemplarisch sowohl auf reduzierte, als auch genomskalige Modelle von *E. coli* angewendet. Dabei konnte eine Vielzahl an Metaboliten identifiziert werden, für deren Synthese ein potentieller thermodynamischer Vorteil besteht, wenn diese in einer Gemeinschaft produziert werden. Die grundsätzlichen theoretischen Annahmen von ASTHERISC und dem damit vorhersagbaren thermodynamischen Gemeinschaftsvorteil wurden zusätzlich an minimalen Stoffwechselmodellen sowohl analytisch als auch in einem kinetischen Modell überprüft.

Alle drei neuen Methoden können somit wertvolle Beiträge zur grundlegenden Theorie und zur biotechnologischen Anwendung constraint-basierter Modellierungstechniken liefern.

Summary

Constraint-based modeling makes it possible to mathematically represent, analyze, and predict the metabolism of organisms. In the last 4 decades, a variety of constraint-based modeling methods have been developed. These methods make use of a wide range of biochemical and physical data.

In this work, three new constraint-based modeling methods are presented: The first new method is sMOMENT, which allows enzymatic constraints to be integrated into constraint-based models with only a minimal expansion of the model itself. Thus, the size of metabolic models with enzymatic constraints can be greatly reduced compared to previous methods with enzymatic constraints. This advantage, in turn, can significantly reduce the runtime of computationally expensive analyses and methods on these models. As a reference implementation of sMOMENT, the AutoPACMEN software package has been developed, which is the first such package that allows the user to simultaneously collect relevant enzymatic data, build an enzymatically constrained model, and calibrate the model. As an example, AutoPACMEN and sMOMENT have been applied to a genome-scale model of *Escherichia coli*. The resulting model can successfully predict the overflow metabolism of *E. coli* under both aerobic and anaerobic conditions. The second new method is TCOSA. It uses the notion of network-wide thermodynamic driving forces to analyze the relevance of redundant cofactor pairs. By applying TCOSA to a genome-scale model of *E. coli* and its NAD(P)(H) reaction specificities, it was shown that the evolution of these specificities is apparently strongly shaped by thermodynamic constraints. In addition, this model could be used to predict qualitative *in vivo* trends in the redox states of NAD(H) and NADP(H). Furthermore, it was shown that the presence of more than two cofactor pairs is thermodynamically advantageous only if the standard redox potential of a third cofactor pair is lowered with respect to NAD(P)(H). The third new method is ASTHERISC. It allows the prediction of thermodynamically favorable product synthesis pathways with multi-strain communities of a species compared to single-strain solutions. ASTHERISC is also based on network-wide thermodynamic driving forces and has been applied to reduced and genome-scale models of *E. coli*. This analysis identified numerous metabolites for which there is a thermodynamic community advantage. In addition, the basic theoretical assumptions of ASTHERISC and the associated predicted thermodynamic community advantage were tested on minimal metabolic models by analytical means and a kinetic model.

Thus, all three new methods can make fruitful contributions to the basic theory and biotechnological application of constraint-based modeling techniques.

Abkürzungsverzeichnis

Symbol	Bedeutung
AutoPACMEN	Automatische Integration von Protein-Allokations-Nebenbedingungen in metabolischen Netzwerken (englisch <i>Automatic Integration of Protein Allocation Constraints in Metabolic Networks</i>)
ASTHERISC	Algorithmische Suche thermodynamischer Vorteile in Einzelspezies-Gemeinschaften (englisch <i>Algorithmic Search of Thermodynamic advantages in Single-species Communities</i>)
CBM	Constraint-basierte Modellierung
CNA	<i>CellNetAnalyzer</i>
DNA	Desoxyribonukleinsäure (englisch <i>deoxyribonucleic acid</i>)
FBA	Flussbilanzanalyse
FVA	Flussvariabilitätsanalyse
gDW	Gramm Trockenmasse (englisch <i>gram dry weight</i>)
GECKO	Genomsskalige Modellerweiterung mittels enzymatische Nebenbedingungen, welche kinetische und „omische“ Daten beachten (englisch <i>Genome-scale Model Enhancement with Enzymatic Constraints Accounting for Kinetic and Omics Data</i>)
MCS	Minimale Schnittmenge im Sinne des Konzepts der <i>Minimal Cut Sets</i>
MDF	Maximal-minimale thermodynamische Triebkraft (englisch <i>max-min driving force</i>)
MILP	Gemischt-ganzzahliges lineares Programm (englisch <i>mixed-integer linear program</i>)
RBA	Ressourcenbilanzanalyse (englisch <i>Resource Balance Analysis</i>)
sMOMENT	Kurz formulierte metabolische Modellierung mit Enzymkinetiken (englisch <i>Short Metabolic Modeling with Enzyme Kinetics</i>)
SubMDF	Maximal-minimale thermodynamische Triebkraft einer ausgewählten Teilmenge an Reaktionen eines metabolischen Modells (englisch <i>subnetwork max-min driving force</i>)
TCOSA	Thermodynamik-basierte Cofaktor-Austausch-Analyse (englisch <i>Thermodynamics-based COfactor Swapping Analysis</i>)
u.d.N.	Unter den Nebenbedingungen

Inhaltsverzeichnis

Zusammenfassung.....	I
Summary.....	II
Abkürzungsverzeichnis.....	III
1 Einleitung.....	1
2 Grundlagen.....	4
2.1 Modellierung von Stoffwechseln.....	4
2.1.1 Modelle metabolischer Netze.....	4
2.1.2 Kinetische Modellierung von Reaktionsflüssen.....	5
2.1.3 Constraint-basierte Modellierung.....	7
2.1.4 Flussbilanz- und Flussvariabilitätsanalyse.....	8
2.1.5 Modellierung mit enzymatischen Nebenbedingungen.....	9
2.1.6 Modellierung mit thermodynamischen Nebenbedingungen.....	13
2.1.7 Minimal Cut Sets.....	16
2.1.8 Modellierung mikrobieller Gemeinschaften.....	17
2.2 <i>Escherichia coli</i> und dessen Modellierung.....	19
2.2.1 Allgemeines und Stoffwechsel.....	19
2.2.2 Constraint-basierte Modelle von <i>E. coli</i>	20
2.3 Wichtige informatische Projekte zur Stoffwechselmodellierung.....	20
2.3.1 Datenbanken und Formate metabolischer Modelle.....	20
2.3.2 Software zur kinetischen Modellierung.....	20
2.3.3 Software zur constraint-basierten Modellierung.....	21
3 Automatische Generierung constraint-basierter metabolischer Modelle mit effizient formulierten enzymatischen Nebenbedingungen.....	22
3.1 Motivation.....	22
3.2 Die sMOMENT-Methode.....	22
3.3 AutoPACMEN.....	26
3.4 Anwendung von AutoPACMEN an einem <i>E. coli</i> -Modell.....	29
3.4.1 Erstellung & Kalibrierung des Modells.....	29
3.4.2 Resultate mit dem kalibrierten Modell.....	32
3.5 Diskussion.....	38
4 Erklärung von NAD(P)(H)-Spezifitäten metabolischer Reaktionen mittels thermodynamischer Modellierung.....	41
4.1 Motivation.....	41
4.2 Das TCOSA-Framework.....	42
4.3 Anwendung von TCOSA auf <i>E. coli</i> und Analyse der Resultate.....	48
4.4 Diskussion.....	64
5 Design mikrobieller Gemeinschaften zur Maximierung der thermodynamischen Triebkraft von Produktsynthesen.....	69
5.1 Motivation.....	69
5.2 Darstellung der Prinzipien von ASTHERISC anhand eines kinetischen Modells.....	73
5.3 Die ASTHERISC-Methode.....	78
5.4 Implementierung von ASTHERISC.....	84
5.5 Anwendung von ASTHERISC an Gemeinschafts-Modellen von <i>E. coli</i> -Stämmen.....	86
5.6 Diskussion.....	96
6 Zusammenfassende Diskussion und Ausblick.....	100
7 Anhang.....	104
8 Literaturverzeichnis.....	110
9 Liste der referenzierten elektronischen Repositories.....	116
10 Liste der Veröffentlichungen in Bezug auf diese Arbeit.....	117

1 Einleitung

Die gesamte lebende Natur basiert auf Stoffwechselprozessen. Diese Prozesse beruhen auf, zumeist durch Enzyme katalysierte, biochemische Reaktionen, bei deren Ablauf Stoffumwandlungen stattfinden. Die sich daraus ergebenden Produkte und die bei ihrer Herstellung freigesetzte Energie werden von den Organismen unter anderem zur Selbsterhaltung und Fortpflanzung genutzt.

Der wissenschaftliche Drang, die Reaktionen des Stoffwechsels zu charakterisieren und zu verstehen, führte schon vor mehr als 100 Jahren zur Schaffung mathematischer Beschreibungen auf diesem Gebiet. Sehr bekannt ist in diesem Zusammenhang die Michaelis-Menten-Kinetik [1]. In ihrer einfachsten Form beschreibt sie die Abhängigkeit der Flussrate einer biochemischen Reaktion von einem ihrer katalysierenden Enzyme. Über enzymkinetische Einflüsse hinaus können Reaktionen auch im Hinblick ihre thermodynamischen Beschränkungen untersucht werden. Deren Definitionen basieren auf der, ebenfalls wohlbekanntem, Gibbs-Energie [2].

Über einzelne Reaktionen hinaus wurden, als Grundlage für Berechnungen, Modelle ganzer Netzwerke von Stoffwechselwegen entwickelt. Diese basierten zunächst auf der Entdeckung und Charakterisierung wichtiger Stoffwechselwege durch biochemische Analysen von Organismen, bei denen die einzelnen spezifischen Stoffumwandlungen aufgeklärt wurden. Später, mit der Entwicklung der DNA-Sequenzierung [3, 4], wurde es zusätzlich möglich, allein auf der Basis des vorhandenen Wissens über biochemische Reaktionen sowie der Genomsequenz eines Organismus, automatisch auf alle bislang charakterisierten Stoffwechselwege in einem Organismus zu schließen [5]. Heutzutage gibt es solche genombasierten Stoffwechselnetzwerke für Tausende verschiedener Arten [6].

Unter anderem zur mathematisierten Beschreibung und Analyse von Stoffwechselnetzwerken wurde eine Vielzahl theoretischer Methoden entwickelt, welche oftmals auf den Prinzipien der sogenannten constraint-basierten metabolischen Modellierung basieren. Diese beruhen auf dem mathematischen Konzept der konvexen Programmierung und benötigen, um den Stoffwechsel eines Organismus zu modellieren, in ihrer einfachsten Form nur die Stöchiometrien der modellierten Reaktionen sowie konvexe Nebenbedingungen (englisch *constraints*) als Eingabe. Hierbei wird zumeist aufgrund ihrer hohen Recheneffizienz speziell die Unterkategorie der linearen Programmierung mit linearen Termen verwendet. Mittels constraint-basierter Methoden lassen sich etwa mögliche Flussraten einer Reaktion in einem metabolischen Modell berechnen. Dies ist auch ein Ziel der wichtigen Methode der Flussbilanzanalyse [7, 8], sowie der darauf aufbauenden Flussvariabilitätsanalyse [9]. Unter anderem zum Bereitstellen dieser Methoden entwickelte sich eine Vielzahl an Programmpaketen, darunter die für diese Arbeit speziell relevanten Projekte COBRApy [10] sowie *CellNetAnalyzer* [11].

Neben den grundlegenden constraint-basierten Methoden wurden solche entwickelt, welche die Integration zusätzlicher biochemischer Daten ermöglichen. Zu diesen Daten gehören etwa enzymkinetische Parameter, wie sie in der Michaelis-Menten-Kinetik verwendet werden. Während diese Kinetik insgesamt nichtlinear ist, wurden lineare Teile dieser in constraint-basierte Methoden integriert. Wichtige Ansätze, die für diesen Zweck entwickelt wurden, sind MOMENT [12] und GECKO [13]. Eine weitere Kategorie der verwendeten biochemischen Daten sind

thermodynamische Parameter. Diese Werte ermöglichen die Eingrenzung des Lösungsraumes innerhalb der constraint-basierten Methoden auf Reaktionen, die thermodynamisch möglich sind, wenn sie gemeinsam aktiv sind. Bekannte Methoden, die solche thermodynamische Nebenbedingungen berücksichtigen, sind TFBA [14] und OptMDFpathway [15].

Eine weitere Möglichkeit zur Erweiterung constraint-basierter Methoden besteht darin, diese Methoden auf Gemeinschaften von Organismen anzuwenden. Solche Gemeinschaften kommen nicht nur in der Natur vor, sondern können auch rational zusammengestellt werden, was sie für biotechnologische Anwendungen besonders interessant macht [16–20]. Zwei wichtige Ansätze für die Modellierung in diesem Zusammenhang sind SteadyCom [21] und RedCom [22].

Die constraint-basierte Modellierung von Stoffwechselnetzen kann nicht nur der Beschreibung und direkten Analyse dienen, sondern unter anderem auch der Vorhersage von Steuerungsmöglichkeiten des Stoffwechsels, dem sogenannten Metabolic Engineering [23]. Ziel dessen ist es, die Leistung eines Organismus durch gezielte gentechnische Veränderungen zu verbessern. Ein gängiger gentechnischer Ansatz *in vivo* hierfür ist das Ausschalten bzw. der Knock-Out von Genen, wodurch assoziierte „störende“ Enzyme nicht mehr produziert werden können, und die durch die Enzyme katalysierten Reaktionen nicht mehr ablaufen können. Eine andere Methode ist der Knock-In, bei dem vorteilhafte Reaktionen eingeführt werden oder *in vivo* inaktive Gene aktiviert werden. Diese Eingriffe zielen häufig darauf ab, die Produktion wertvoller Industrieprodukte in einem Organismus zu verbessern. Diese Verbesserung kann z.B. darin bestehen, das Verhältnis zwischen Produkt und Substrat zu erhöhen oder die Produktionsrate in einer bestimmten Zeitspanne zu steigern.

Im Sinne des Metabolic Engineering wurden constraint-basierte Techniken entwickelt, die eine direkte Vorhersage gentechnischer Eingriffe im Rahmen eines rationalen Stoffwechselfdesigns ermöglichen. Insbesondere Methoden wie OptKnock [24], RobustKnock [25] und OptCouple [26] sind von großer Bedeutung. Darüber hinaus zeichnet sich die Methode der Minimal Cut Sets (MCS) als der flexibelste Ansatz in Bezug auf einsetzbare Nebenbedingungen aus [27]. MCS wurden bereits erfolgreich *in vivo* eingesetzt, z.B. um die Synthese von Produkten zu optimieren [28]. Oftmals wird im Rahmen dieser Optimierung auch das Konzept der starken Wachstumskoppelung (englisch *growth coupling*) angewandt [27, 29], sodass eine Zelle nach der MCS-errechneten Intervention nur noch dann wachsen kann, wenn das Produkt mitsynthetisiert wird.

Die Methoden der constraint-basierten Modellierung werden außerdem nicht nur für die Suche nach konkreten Lösungen für eine optimierte Produktion mit Organismen verwendet. So dienen diese beispielsweise auch zur Entdeckung möglicher Prinzipien hinter der Evolution von Stoffwechselwegen [30–32].

Die vorliegende Arbeit basiert auf den genannten Prinzipien und Zielen der constraint-basierten Modellierung. Sie versucht, diese Art der Modellierung durch drei Entwicklungen zu erweitern:

1. In Kapitel 3 wird die neu entwickelte Methode sMOMENT vorgestellt, die im Vergleich zu den früheren Methoden MOMENT sowie GECKO eine deutlich effizientere Methode zum Einbau enzymatischer Nebenbedingungen in constraint-basierte Stoffwechselmodelle darstellt. Darüber hinaus wird die Implementierung dieser Methode, AutoPACMEN, zusammen mit ihrer Anwendung auf ein genomskaliges Modell von *E. coli* dargelegt.

2. In Kapitel 4 wird das TCOSA-Framework vorgestellt und erläutert, wie es auf die NAD(P)(H)-Spezifitäten in einem genomskaligen Modell von *E. coli* angewendet wird. Mittels TCOSA kann generell der thermodynamische Einfluss der Spezifität von Reaktionen bei redundanten Cofaktorrauen untersucht werden. Dies wird durch das Konzept der netzwerkweiten thermodynamischen Triebkräfte erreicht. Wie in diesem Kapitel gezeigt, sind diese Triebkräfte im Anwendungsfalle der *in vivo* vorkommenden NAD(P)(H)-Spezifitäten von *E. coli* nahe oder am theoretischen Optimum. Dies ist gleichzeitig ein starkes Indiz dafür, dass die Optimierung der genannten Triebkräfte die Evolution der NAD(P)(H)-Spezifitäten in *E. coli* stark beeinflusst hat. Weitere im Kapitel dargestellte Untersuchungen zeigten zudem, dass auch qualitative Trends des Redoxzustandes von NAD(H) sowie NADP(H) korrekt vorhergesagt werden können, und dass die genannten Ergebnisse auch unter unterschiedlichen getesteten Bedingungen resultieren können.
3. In Kapitel 5 wird der ASTHERISC-Algorithmus vorgestellt. Dieser Algorithmus konzentriert sich auf constraint-basierte Stoffwechselmodelle mikrobieller Gemeinschaften, die aus Stämmen einer einzigen Art bestehen. ASTHERISC kann alle Fälle aufzeigen, in denen eine solche Gemeinschaft einen thermodynamischen Vorteil bei der Produktsynthese gegenüber einem einzelnen Stamm hat. Als Beispielfall wird ASTHERISC in diesem Kapitel auf entsprechende Gemeinschaftsmodelle von *E. coli* angewendet. Hierbei wurden dutzende Fälle identifiziert, bei denen eine Multi-Stamm-Gemeinschaft den genannten Vorteil bei Produktsynthesen erzielt.

Um das Verständnis dieser neu entwickelten Methoden zu vereinfachen, werden in Kapitel 2 die für diese Arbeit relevanten biologischen, mathematischen und informatischen Grundlagen dieser Methoden dargelegt. Anschließend werden in Kapitel 6 die Ergebnisse zusammengefasst und ein Ausblick auf mögliche zukünftige Vorhaben gegeben.

Ein Großteil der in dieser Arbeit dargestellten Resultate wurde in den folgenden Publikationen veröffentlicht, welche hier nicht weiter zitiert werden:

- Bekiaris & Klamt, 2020. Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*, 21(1), 1-13.
- Bekiaris & Klamt, 2021. Designing microbial communities to maximize the thermodynamic driving force for the production of chemicals. *PloS Computational Biology*, 17(6), e1009093.
- Bekiaris & Klamt, 2023. Network-wide thermodynamic constraints shape NAD(P)H cofactor specificity of biochemical reactions. *Nature Communications* 14, 4660.

Weitere Publikationen unter Mitwirkung des Autors, die in einem entfernteren Zusammenhang zu dieser Arbeit stehen, sind:

- Thiele, von Kamp, Bekiaris, Schneider, & Klamt, 2022. CNApy: a *CellNetAnalyzer* GUI in Python for analyzing and designing metabolic networks. *Bioinformatics*, 38(5), 1467-1469.
- Schneider, Bekiaris, von Kamp & Klamt, 2022. StrainDesign: a comprehensive Python package for computational design of metabolic networks. *Bioinformatics*, 38(21), 4981-4983.

2 Grundlagen

2.1 Modellierung von Stoffwechseln

2.1.1 Modelle metabolischer Netze

Um ein Stoffwechselnetz mathematisch analysieren zu können, muss dieses Netz in ein Modell überführt werden. Ein solches Modell besteht aus folgenden zentralen Komponenten:

1. **Metabolite:** Dies sind alle Stoffe, die am modellierten Stoffwechsel beteiligt sind. Enthält der modellierte Organismus mehrere Kompartimente, etwa ein Cytoplasma und ein Periplasma, kann ein chemisch identischer Metabolit im Modell mehrfach vorkommen. In diesem Falle existiert für jedes Kompartiment eine eigene Version des Metaboliten. Einige Modellierungstechniken unterscheiden zwischen internen und externen Metaboliten, wobei letztere ohne eine entsprechende Austauschreaktion erzeugt werden können (siehe nächsten Absatz). Alle im Rahmen dieser Arbeit betrachteten constraint-basierten Modelle enthalten ausschließlich interne Metabolite.
2. **Reaktionen:** Diese stellen alle modellierten metabolischen Stoffumwandlungen dar. Jede Reaktion benötigt als Information die Identität sowie Stöchiometrie der assoziierten Ausgangsstoffe und der Produkte. Zusätzlich zu den Stoffwechselreaktionen enthalten typische Stoffwechselmodelle auch „Pseudoreaktionen“, die mechanistisch als Reaktion nicht in dem zu modellierenden Organismus ablaufen, aber für die Modellierung notwendig sind. Eine davon ist die Biomasse- oder Wachstumsreaktion, bei der die Stöchiometrie essentieller Stoffe wie Aminosäuren oder Lipide als Substrate, und manchmal auch als Produkte, der Biomassebildung eingehen. Darüber hinaus enthalten viele Modelle eine Erhaltungsreaktion, bei der eine bestimmte Menge ATP zu ADP verbraucht werden muss, um den vom Wachstum unabhängigen Energieverbrauch zur Erhaltung der Zelle zu modellieren. Zudem existiert eine Reihe an Austauschreaktionen, bei denen es nur ein Substrat oder ein Produkt gibt. Diese Substrate bzw. Produkte stellen Metabolite dar, die der Organismus aus seiner Umgebung aufnimmt oder in diese ausscheidet.
3. **Enzyme:** Diese stellen diejenigen Enzyme dar, welche die im modellierten Stoffwechsel enthaltenen enzymabhängigen Reaktionen katalysieren. Jede Reaktion kann mit den Enzymen, die für die Katalyse dieser Reaktion verantwortlich sind, durch Enzym-Reaktions-Regeln verknüpft werden. Diese Regeln enthalten die logischen Operatoren UND sowie ODER, die mit den Namen der Enzyme arbeiten. Darüber hinaus gibt es Klammern, welche die Präzedenz der verwendeten Verknüpfungen angeben und eine Verschachtelung der Regeln ermöglichen. In Ergänzung oder alternativ zu den Enzymen lassen sich in einem Modell auch die Gene der Enzyme einbauen. Jedes Gen steht dabei für einen enzymkodierenden Teil des Genoms. Entsprechend ändern sich die genannten Regeln dann zu Gen-Reaktions- oder Gen-Enzym-Reaktions-Regeln.

Während die Definition von Enzymen für ein metabolisches Modell optional ist, ist die Definition von Metaboliten und Reaktionen essentiell. Die beiden letzteren lassen sich in einer stöchiometrischen Matrix N zusammenfassen, die für alle nachfolgend beschriebenen Methoden

notwendig ist. N hat so viele (n) Spalten wie Reaktionen und (m) Zeilen wie Metabolite. Jeder Eintrag dieser Matrix enthält die Stöchiometrie des Metaboliten in der jeweiligen Reaktion. Ist der Eintrag negativ, wird der Metabolit in der Reaktion als Substrat verbraucht, ist er positiv, wird der Metabolit als Produkt hergestellt.

Es existiert eine Vielzahl an Programmpaketen zur computergestützten Erstellung metabolischer Modelle eines Zielorganismus. Wichtige Beispiele sind CarveMe [33], merlin [34] sowie Model SEED [35]. Das Grundprinzip ist bei all diesen Softwarepaketen gleich und wird z.B. in [5] näher beschrieben: Die aus der Genomsequenzierung des Zielorganismus resultierende Genomsequenz wird zunächst zur Bestimmung der Gene verwendet. Aus den Gensequenzen werden dann die von diesen Genen kodierten Enzyme bestimmt. Aus der Menge dieser Enzyme kann dann auf die Reaktionen im Zielorganismus geschlossen werden, die durch diese Enzyme katalysiert werden. Darüber hinaus werden die oben erwähnten Pseudoreaktionen passend eingefügt.

2.1.2 Kinetische Modellierung von Reaktionsflüssen

Ist eine biochemische Reaktion aktiv, so läuft sie mit einem bestimmten Reaktionsfluss ab. Die in der constraint-basierten Modellierung oft verwendete Einheit des Reaktionsflusses ist $\text{mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$, wobei „gDW“ für „Gramm Trockenmasse“ aus Englisch „gram dry weight“ steht. Eine Ausnahme bildet die Biomasse-Reaktion, deren Fluss in der Regel die Einheit h^{-1} hat.

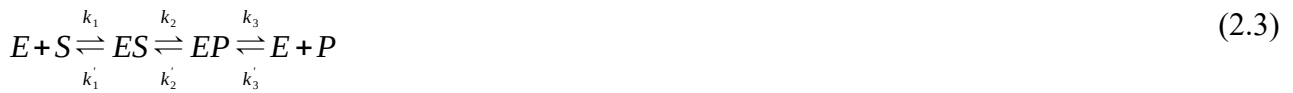
Eine weitverbreitete mathematische Beschreibung des Flusses bzw. der Rate enzymatisch katalysierter Reaktionen stellt die irreversible Michaelis-Menten-Kinetik [1] dar. Reaktionen, die dieser Kinetik in der Realität folgen können, werden durch folgenden Mechanismus (hier für eine beispielhafte Reaktion $S \rightarrow P$) beschrieben, wenn sie durch thermodynamische Effekte nicht in ihrem Fluss eingeschränkt sind:



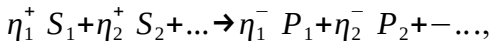
E steht hierbei für das die Reaktion katalysierende Enzym, ES für einen sich im Laufe der Reaktion bildenden temporären Komplex aus Substrat sowie Enzym, und k_1 , k_{-1} sowie k_2 sind die Geschwindigkeitskonstanten der jeweils angegebenen Teilreaktion. Die sich nun hieraus ergebende wohlbekanntere irreversible Michaelis-Menten Kinetik ist unter den in [1] getroffenen Annahmen (wie etwa $[S] \gg [E]$ und dass ein Fließgleichgewicht (siehe weiter unten) in dem beschriebenen Reaktionsmechanismus vorliegt)

$$r = [E] \cdot k_{cat}^+ \cdot \frac{[S]}{[S] + K_S} \quad (2.2)$$

r ist der Fluss der Reaktion, $[E]$ ist die Konzentration des reaktionskatalysierenden Enzyms, k_{cat}^+ die Wechselzahl des Enzyms in der angegebenen Vorwärtsrichtung der Reaktion, d.h. die Zahl der von diesem Enzym katalysierten Substratumsetzungen pro Zeiteinheit, $[S]$ die Substratkonzentration und K_S die Michaeliskonstante des Enzyms in Bezug auf diese Reaktion und das Substrat, d.h. die Substratkonzentration, bei der der Reaktionsfluss die Hälfte seines durch den Term $[E] \cdot k_{cat}^+$ beschränkten Maximums erreichen kann. Die irreversible Michaelis-Menten-Kinetik kann auch, wie in [36] gezeigt, für reversible Reaktionen mit dem Mechanismus



erweitert werden. Hier wird also explizit auch die Bildung des Komplexes aus Enzym und Produkt, EP , mit entsprechenden weiteren Geschwindigkeitskonstanten k_2' , k_3 sowie k_3' betrachtet. Dieser Mechanismus wird in [36] zugrundegelegt, um zu einer erweiterten und für mehr Fälle anwendbaren Formulierung von Gleichung (2.2) zu gelangen. Diese Formulierung kann zusätzlich, als weitere Generalisierung, für eine beliebige Anzahl von Substraten und Produkten der Formulierung



wobei η_i^+ bzw. η_j^- für die absolute Stöchiometrie des jeweiligen Substrats S bzw. Produkts P in der Reaktion steht, verwendet werden und kann zudem den Einfluss thermodynamischer Effekte auf den Reaktionsablauf einbeziehen. Konkret lautet diese in [36] entwickelte Formulierung

$$r = V^+ \cdot \kappa \cdot \gamma. \quad (2.4)$$

Der Fluss r kann hierbei sowohl positiv (Ablauf der Vorwärtsrichtung) als auch negativ (Ablauf der Rückwärtsrichtung) werden. V^+ steht für den maximal möglichen Reaktionsfluss in eine angegebene Richtung und ist definiert als

$$V^+ = [E] \cdot k_{cat}^+ \quad (2.5)$$

mit $[E]$ als Konzentration des Enzyms oder Enzymkomplexes.

κ steht für den allgemeinen Sättigungsterm und ist definiert als

$$\kappa = \frac{\prod_i ([S_i]/K_{S,i})^{\eta_i^+}}{1 + \prod_i ([S_i]/K_{S,i})^{\eta_i^+} + \prod_j ([P_j]/K_{P,j})^{\eta_j^-}} \quad (2.6)$$

mit $[S_i]$ für die Konzentration des i -ten Substrats der Reaktion, $[P_j]$ für die Konzentration des j -ten Produkts und $K_{S,i}$ bzw. $K_{P,j}$ für die Michaeliskonstante des Enzyms zum jeweiligen Substrat bzw. Produkt steht.

γ stellt die thermodynamischen Einflüsse dar, die den maximal möglichen Reaktionsfluss begrenzen und ist definiert als

$$\gamma = 1 - e^{\frac{\Delta_r G'}{R \cdot T}} \quad (2.7)$$

mit $\Delta_r G'$ als die Änderung der Gibbschen Reaktionsenergie, welche im Abschnitt 2.1.6 näher erläutert wird, R als Gaskonstante, sowie T als Temperatur.

Sowohl κ als auch γ können unter physiologisch möglichen Definitionsbereichen ihrer Parameter bei der Vorwärtsrichtung einer aktiven Reaktion (mit positiven Konzentrationen sowie negativem $\Delta_r G'$) nur Werte im Intervall $[0;1]$ annehmen. Somit können sie einen Reaktionsfluss einer ablaufenden Reaktion nicht erhöhen, sondern allenfalls den durch V^+ gegebenen maximalen Fluss erniedrigen.

Mathematische Beschreibungen einzelner Reaktionsflüsse, wie die dargestellte reversible Michaelis-Menten-Kinetik (Gleichung (2.4)), können in kinetische Modelle ganzer Reaktionsnetzwerke überführt werden, wie z.B. in [37] näher beschrieben. Diese Modelle stellen Differentialgleichungssysteme dar, die den zeitlichen Ablauf von Reaktionsflüssen in einem metabolischen Netzwerk beschreiben können. Im Allgemeinen lassen sich solche Systeme, analog zur allgemeinen Formulierung in [38], durch die zentrale Differentialgleichung

$$\frac{d\mathbf{c}(t)}{dt} = \mathbf{N} \cdot \mathbf{r}(t) \quad (2.8)$$

beschreiben. Hierbei ist \mathbf{N} konstant und steht für die schon erwähnte stöchiometrische Matrix $\mathbf{N} \in \mathbb{R}^{m \times n}$ mit m Metaboliten und n Reaktionen, $\mathbf{c}(t)$ für einen zeitabhängigen Vektor der Länge m , der die von der Zeit t abhängigen Metabolitkonzentrationen beinhaltet, und $\mathbf{r}(t)$ für einen zeitabhängigen Flussvektor der Länge n , welcher wiederum über die Gleichung

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

bestimmt wird. \mathbf{f} ist hierbei die Funktion, welche die Reaktionsflüsse in $\mathbf{r}(t)$ in Abhängigkeit von $\mathbf{c}(t)$ und einem Vektor \mathbf{p} , in welchem alle für die Flüsse relevanten Parameter enthalten sind, bestimmt.

Eine typische und für diese Arbeit relevante Berechnungsmethode der kinetischen Modellierung ist die Simulation in Form einer Zeitreihe. Dabei werden zunächst für jede Zustandsgröße Anfangswerte gesetzt [37]. Danach wird die Zeit in diskreten Schritten fortgesetzt. In jedem dieser Schritte werden die Zustandsgrößen durch numerische Integration der Differentialgleichungen des kinetischen Modells aktualisiert, ausgehend von den zum vorherigen Zeitpunkt berechneten Werten des Modells. Mithilfe von Zeitreihen lässt sich beispielsweise das dynamische Verhalten eines Modells betrachten.

Neben den Zeitreihen gibt es eine Vielzahl weiterer Berechnungsmethoden und Analysen mit kinetischen Modellen [37], darunter z.B. allgemeine Sensitivitätsanalysen für die Auswirkung der Änderung eines Parameters auf die Lösungen im Modell. Letzteres ist auch eine zentrale Grundidee der metabolischen Kontrollanalyse (englisch *metabolic control analysis*), bei der u.a. die Auswirkungen kleinerer Parameteränderungen eines metabolischen Modells untersucht werden [39].

Trotz der potentiell genaueren Vorhersagen hat die kinetische Modellierung einige wesentliche Nachteile. So ist die Optimierung mit kinetischen Modellen auf besonders rechenintensive heuristische Algorithmen angewiesen. Dies wiederum beschränkt in der Praxis die effektive Größe für die Optimierung verwendbarer kinetischer Modelle auf einige Dutzend bis hundert Reaktionen. Darüber hinaus sind kinetische Modelle von einer großen Anzahl biochemischer Parameter abhängig (\mathbf{p} in Gleichung (2.8)), von denen viele nur mit großer Unsicherheit oder gar nicht gemessen wurden [40].

2.1.3 Constraint-basierte Modellierung

Eine weit verbreitete Alternative zur kinetischen Modellierung eines Stoffwechsels ist die constraint-basierte Modellierung, kurz CBM. Diese basiert auf Nebenbedingungen (englisch

constraints), die in Form konvexer Terme formuliert werden müssen. Die vorliegende Arbeit konzentriert sich dabei ausschließlich auf die constraint-basierte Modellierung mit (konvexen) linearen Termen, die zuerst in [8] konkret für metabolische Netzwerke angewendet wurde.

Die Methoden der constraint-basierten Modellierung nutzen das Konzept des Fließgleichgewichts (englisch *steady state*) aller modellierten Metabolitkonzentrationen [37]. Dies beruht auf der Grundannahme, dass, etwa im Vergleich zur Wachstumsrate einer Zelle, die Umsatzraten von Metaboliten um Größenordnungen schneller sind und daher nur zeitlich konstante Konzentrationen ein stabiles Wachstum einer Zelle ermöglichen [41]. Dieses Fließgleichgewicht kann, unter Nutzung der Definitionen aus Gleichung (2.7), formalisiert werden als

$$\frac{d\mathbf{c}(t)}{dt} = \mathbf{N} \cdot \mathbf{r}(t) = \mathbf{0}.$$

Bei der CBM ist man aber nicht am Vektor $\mathbf{c}(t)$ interessiert und betrachtet $\mathbf{r}(t)$ als von der Zeit unabhängigen Vektor \mathbf{r} , sodass im Rahmen der CBM das Fließgleichgewicht als

$$\mathbf{N} \cdot \mathbf{r} = \mathbf{0}, \tag{2.9}$$

formuliert wird, womit weiterhin das genannte Fließgleichgewicht gefordert wird. Jeder einzelne Reaktionsfluss kann zudem eingeschränkt werden mit Nebenbedingungen der Form

$$\alpha_i \leq r_i \leq \beta_i, \tag{2.10}$$

wobei α_i für den minimal und β_i für den maximal erlaubten Fluss der i -ten Reaktion steht. Gilt $\alpha_i \geq 0 \wedge \beta_i > 0$ ist die Reaktion irreversibel und kann nur in die Vorwärtsrichtung ablaufen. Analog kann bei $\alpha_i < 0 \wedge \beta_i \leq 0$ die Reaktion nur in die Rückwärtsrichtung laufen. Gilt $\alpha_i < 0 \wedge \beta_i > 0$, so ist die Reaktion reversibel und kann entweder in die Vorwärts- oder Rückwärtsrichtung ablaufen.

Ferner können freie weitere lineare Nebenbedingungen der Form

$$\mathbf{A} \cdot \mathbf{r} \leq \mathbf{b} \tag{2.11}$$

eingeführt werden. \mathbf{A} ist eine Matrix, $\mathbf{A} \in \mathbb{R}^{q \times n}$, mit q als Anzahl der weiteren Nebenbedingungen. Ein typisches Beispiel für solche weiteren Nebenbedingungen ist die Definition einer minimalen Ausbeute wie z.B. ein minimales Verhältnis an Substrat- zu Produktsynthese.

2.1.4 Flussbilanz- und Flussvariabilitätsanalyse

Eine wesentliche Optimierungstechnik auf Basis der CBM ist die *Flussbilanzanalyse* (englisch *Flux Balance Analysis*) oder kurz FBA [7, 42]. Bei einer FBA wird, unter Nutzung der Nebenbedingungen der CBM, eine Flussrate per

$$\underset{\mathbf{r}}{\text{maximiere}} \quad \mathbf{c}^T \cdot \mathbf{r} \tag{2.12}$$

u. d. N. (2.9) bis (2.11)

optimiert. Hierbei ist \mathbf{c} ein Vektor der Länge n und weist jedem Reaktionsfluss einen Koeffizienten für die Zielfunktion zu. Möchte man alternativ eine Minimierung der Zielfunktion durchführen, so muss man nur das Vorzeichen der Einträge in \mathbf{c} tauschen.

Möchte man für jeden Reaktionswert einzeln den maximal und minimal möglichen Flusswert berechnen, so bietet sich die Flussvariabilitätsanalyse (englisch *Flux Variability Analysis*), kurz FVA, an [9]. Ein häufiges Beispiel ist die FVA nach der Optimierung der Zielfunktion durch eine FBA. Hier kann nach der FBA die Nebenbedingung

$$\mathbf{c}^T \cdot \mathbf{r} \geq \omega \quad (2.13)$$

eingeführt werden, wobei ω für das zuvor ermittelte Optimum der Zielfunktion steht. Die FVA funktioniert jedoch optional auch ohne diese zusätzliche Nebenbedingung.

Bei der FVA wird nun für jede einzelne Reaktionsrate eine Maximierung sowie Minimierung per

$$\begin{aligned} &\underset{\mathbf{r}}{\text{minimiere}} \ r_i && (2.14) \\ &\text{u. d. N. (2.9) bis (2.11), (2.13)} \end{aligned}$$

sowie

$$\begin{aligned} &\underset{\mathbf{r}}{\text{maximiere}} \ r_i && (2.15) \\ &\text{u. d. N. (2.9) bis (2.11), (2.13)} \end{aligned}$$

durchgeführt. Als Ergebnis erhält man die maximal und minimal möglichen Flüsse der Reaktionen. Dabei ist zu beachten, dass jeweils nur die Flussgrenzen einer einzelnen Reaktion betrachtet werden. Somit ist nicht gegeben, dass jede mögliche Kombination der von der FVA berechneten Reaktionsflüsse unter den gegebenen Nebenbedingungen möglich ist.

Neben direkt linearen Termen können in der CBM auch Ausbeuten (englisch *yield*), wie z.B. ein Verhältnis von Produktsynthese zu Substratverbrauch, optimiert werden, wie etwa in [43] näher dargelegt. Die Optimierung basiert hierbei auf dem Konzept der linear-fractionalen Programmierung.

Die vorgestellten Methoden der FBA sowie der FVA berücksichtigen nur den Fluss einer Reaktion, nicht aber den in Gleichung (2.6) dargestellten Sättigungsterm und, mit Ausnahme der Reversibilität einer Reaktion, nicht die thermodynamischen Beschränkungen auf den Reaktionsfluss. In den folgenden Abschnitten werden hierzu Modellierungstechniken vorgestellt, mit denen diese Einflüsse zumindest teilweise integriert werden können.

2.1.5 Modellierung mit enzymatischen Nebenbedingungen

Eine Erweiterung der CBM ist die Einbeziehung enzymatischer Nebenbedingungen. Hierfür wird zunächst angenommen, dass sowohl der Sättigungsterm als auch thermodynamische Effekte keinen Einfluss auf den Fluss einer Reaktion haben, d.h. $\kappa=1 \wedge \gamma=1$ in Gleichung (2.4), sodass der Fluss einer Reaktion i gleich V^+ bzw.

$$r_i = k_{cat,i} \cdot [E_i] \quad (2.16)$$

ist. Zur Vereinfachung kann diese Gleichung auch als Ungleichung aufgefasst werden, in welcher der maximal mögliche Reaktionsfluss dargestellt wird, da V^+ immer die maximal mögliche obere Grenze eines Flusses darstellt:

$$r_i \leq k_{cat,i} \cdot [E_i]. \quad (2.17)$$

Da diese Gleichung mit negativen Flussraten biologisch unplausibel wird, müssen, bei allen im Folgenden vorgestellten constraint-basierten Methoden mit enzymatischen Nebenbedingungen, reversible Reaktionen in irreversible zerlegt werden. Während solche Nebenbedingungen, die von der Enzymkonzentration abhängen, leicht in ein constraint-basiertes Modell integriert werden könnten, sind in der Praxis die meisten Konzentrationsbereiche metabolisch aktiver Enzyme unbekannt oder auch potentiell stark variabel. Dennoch muss für diese ungemessenen Enzyme eine obere Schranke für die Gesamtmenge der verwendeten Enzyme definiert werden, um zu verhindern, dass ein Modell unendlich viel Enzym verwenden kann. Eine grundlegende Arbeit zu diesem Zwecke war die FBA-basierte „Flux Balance Analysis with Molecular Crowding“ [44], kurz FBAwMC. Dabei wird aus einem durch das Zellvolumen begrenzten Gesamtpool eine bestimmte Enzymmenge in Abhängigkeit vom Reaktionsfluss entnommen. Die Enzymmenge wurde dabei nicht enzymkinetisch abgeleitet, sondern lediglich aus dem Molekulargewicht des Enzyms und einem allgemeinen Parameter abgeschätzt.

Diese Schätzung wurde durch die Methode „Metabolic Modeling with Enzyme Kinetics“ [12], kurz MOMENT, verfeinert. MOMENT beruht auf den Nebenbedingungen der CBM (Gleichungen (2.9) bis (2.11)). Zum einen jedoch führt MOMENT unter anderem Nebenbedingungen gemäß der Gleichung (2.17) ein, welche die Abhängigkeit der Enzymnutzung vom assoziierten Reaktionsfluss beschreiben. Ferner besteht in MOMENT die entscheidende Nebenbedingung für die obere Schranke an Protein

$$\sum_i [E_i] \cdot W_i \leq \Omega, \quad (2.18)$$

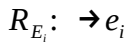
wobei W_i für das Molekulargewicht des jeweiligen Enzyms in $\text{g} \cdot \text{mmol}^{-1}$ und Ω für eine maximal zulässige Gesamtzymmasse pro Biomasse in $\text{g} \cdot \text{gDW}^{-1}$ steht, die im Folgenden kurz als „Gesamt-Enzympool“ bezeichnet wird. In besonderen Fällen, wie etwa Isozymen und Quartärstrukturen, wird Gleichung (2.18) passend modifiziert. Mit einem MOMENT-Modell von *E. coli* konnten sowohl die Wachstumsraten dieses Organismus unter verschiedenen Kohlenstoffquellen, als auch der Überlaufstoffwechsel (englisch *overflow metabolism*) von *E. coli* unter aeroben Bedingungen vorhergesagt werden [12]. Ein Nachteil von MOMENT ist, dass es nicht möglich ist, seine Nebenbedingungen direkt in die stöchiometrische Matrix eines constraint-basierten metabolischen Modells zu integrieren.

Unter anderem zur Lösung des letzteren Problems wurde die Methode „Genome-scale Model Enhancement with Enzymatic Constraints Accounting for Kinetic and Omics Data“ [13], kurz GECKO, entwickelt. GECKO stellt eine, geringfügig erweiterte, Umformulierung von MOMENT dar und basiert wie MOMENT auf den Nebenbedingungen der CBM (Gleichungen (2.9) bis (2.11)). Zunächst erweitert GECKO die Definition des Gesamt-Enzympools Ω zu

$$\Omega = (P_{total} - P_{gemessen}) \cdot \phi \cdot \sigma. \quad (2.19)$$

P_{total} steht für die gesamte Proteinmenge pro Biomasseeinheit in $\text{g} \cdot \text{gDW}^{-1}$, $P_{gemessen}$ für alle schon gemessenen Proteinkonzentrationen (welche nicht in den Gesamt-Enzympool fallen, siehe weiter

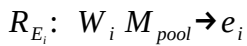
unten), φ für den Anteil der modellierten Enzyme an der Gesamt-Proteinmenge und σ für eine angenommene mittlere Substratsättigung der modellierten Enzyme. Mithilfe der Werte φ sowie σ lässt sich somit der Enzympool zum Zwecke passender Modellergebnisse ändern. Damit nun die Nebenbedingungen von MOMENT direkt in die stöchiometrische Matrix N eingebaut werden können, müssen sie als Teil einer Reaktion eingefügt werden. Hierfür wird zunächst für jedes Enzym der modellierten Reaktionen ein Pseudo-Metabolit e_i eingefügt. Wie e_i erzeugt wird, hängt dabei davon ab, ob Konzentrationsmessungen für das entsprechende Enzym E_i vorliegen: Liegt für E_i eine solche Messung vor, so wird e_i in einer Pseudo-Reaktion der Form



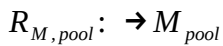
mit den Flussgrenzen

$$0 \leq r_{R_{E_i}} \leq [E_i] \quad (2.20)$$

erzeugt, wobei der Fluss eine Einheit von $\text{mmol} \cdot \text{gDW}^{-1}$ hat. e_i steht somit für die im modellierten Metabolismus gerade aktiv genutzte Konzentration in $\text{mmol} \cdot \text{gDW}^{-1}$. Liegt für E_i hingegen keine Konzentrationsmessung vor, so wird e_i durch irreversible Pseudo-Reaktionen der Form



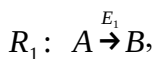
erzeugt, welche irreversibel ist und in ihrem maximal erlaubten Fluss unbeschränkt ist, wobei der Fluss eine Einheit von $\text{mmol} \cdot \text{gDW}^{-1}$ hat. Durch den stöchiometrischen Koeffizienten W_i in R_{E_i} , dem Molekulargewicht von E_i in $\text{g} \cdot \text{mmol}^{-1}$, wird hierbei die Entnahme aus dem Gesamt-Enzympool in $\text{g} \cdot \text{gDW}^{-1}$ in eine Konzentration $[E_i]$ umgewandelt. Beschränkt ist die Reaktion durch den Maximalwert des Gesamt-Enzympools Ω (Gleichung (2.19)). M_{pool} wiederum wird global aus einer eigenen irreversiblen Pseudo-Reaktion



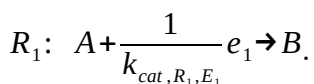
erzeugt, mit den Flussgrenzen

$$0 \leq r_{R_{M, pool}} \leq \Omega, \quad (2.21)$$

wobei der Fluss eine Einheit von $\text{g} \cdot \text{gDW}^{-1}$ hat. Wird nun beispielsweise eine Reaktion $A \rightarrow B$ durch ein einzelnes Enzym E_1 katalysiert, etwa



so wird diese in GECKO, mit ihren originalen Flussgrenzen, wie folgt umgeschrieben:



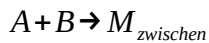
Durch die Stöchiometrie $1/k_{cat, R_1, E_1}$ in der Einheit h wird von e_1 (und damit E_1) die Menge an Enzym entnommen welche zum Erhalt des Flusses gesamten Reaktion in $\text{mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ notwendig ist. Dass diese Herleitung valide ist, lässt sich durch folgendes Beispiel darlegen: Hierfür wird zunächst angenommen, dass R_1 den Fluss $1 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ und E_1 einen k_{cat} von

1 h⁻¹ hat. Mit diesem k_{cat} katalysiert somit E_1 einen Umsatz der Reaktion in einer Stunde. Für einen Fluss von R_1 an $1 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ wird somit 1 mmol an E_1 benötigt. Je höher der k_{cat} wird, umso proportional niedriger wird nun der Bedarf an E_i bei gleichbleibendem Reaktionsfluss, da mehr Reaktionsumsätze in der gleichen Zeit möglich sind. Somit benötigt man, wie in GECKO eingebaut, pro Reaktionsfluss einer Reaktion (in $\text{mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$) tatsächlich $1/k_{\text{cat}, R_i, E_i}$ an $\text{mmol} \cdot \text{gDW}^{-1}$ des Enzyms.

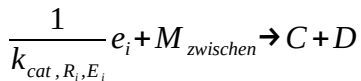
Alle bisher betrachteten Gleichungen behandeln den Fall, dass *eine* Reaktion durch *ein* Enzym katalysiert wird. Die weiteren Möglichkeiten der Enzymverteilung werden in GECKO wie folgt behandelt, wobei diese Fälle beliebig kombiniert werden können:

Im Falle eines promiskuösen Enzyms, also eines Enzyms, welches mehrere verschiedene Reaktionen katalysieren kann, wird der Pseudo-Metabolit e_i , gemäß der schon erwähnten Ergänzung um diesen Pseudo-Metaboliten, in alle betreffenden Reaktionen mit einem entsprechenden von der Reaktion und dem Enzym abhängigen k_{cat, R_i, E_i} eingefügt.

Bei Vorliegen von Isozymen wird die betreffende ursprüngliche Reaktion mehrfach aufgeteilt: Zunächst wird die ursprüngliche Reaktion, bspw. $A + B \rightarrow C + D$, durch eine „Hilfsreaktion“ ersetzt. In dieser Hilfsreaktion wird, anstatt der Produkte der Originalreaktion, ein zwischenliegender Pseudo-Metabolit M_{zwischen} produziert, sodass etwa die Hilfsreaktion im gegebenen Beispiel



lautet. Diese veränderte Reaktion beinhaltet die Flussgrenzen (Gleichung (2.10)) der ursprünglichen Reaktion, sodass diese weiterhin eingehalten bleiben. Nun wird für jedes Isozym mit eigenem e_i der ursprünglichen Reaktion eine eigene M_{zwischen} -konsumierende Reaktion der Form



eingefügt, sodass nun mehrere von den verschiedenen Isozymen abhängige produktsynthetisierende Reaktionen bestehen.

Ist das katalysierende Enzym einer Reaktion tatsächlich ein Enzymkomplex K_j , bestehend aus Unterenzymen E_i , so wird die Reaktion, analog zum beschriebenen Falle mit einem Enzym, um jedes einzelne e_i der Unterenzyme E_i als Substrat mit der Stöchiometrie $\eta_{E_i, K_j} / \eta_{\text{cat}, R_i, E_i}$ erweitert. η_{E_i, K_j} steht hierbei für die interne Stöchiometrie des jeweiligen Unterenzyms innerhalb des gegebenen Enzymkomplexes.

Da GECKO vollständig als Umformulierung von N beschrieben werden kann, sind die Nebenbedingungen und Zielfunktionen der CBM (Gleichungen (2.9) bis (2.11)), der FBA (Gleichung (2.12)) sowie der FVA (Gleichungen (2.13) bis (2.15)) direkt verwendbar. In der initialen Veröffentlichung [13] wurde GECKO zur Erstellung eines um enzymatische Nebenbedingungen erweiterten Modells der Hefe *Saccharomyces cerevisiae* eingesetzt, mit dem auch deren Überlaufmetabolismus reproduziert werden konnte.

2.1.6 Modellierung mit thermodynamischen Nebenbedingungen

Bei der constraint-basierten Modellierung können auch thermodynamische Nebenbedingungen berücksichtigt werden. In der Michaelis-Menten-Kinetik ist der zugrundeliegende thermodynamische Einfluss, wie in Gleichung (2.7) dargestellt, als

$$\gamma = 1 - e^{\frac{\Delta_r G'}{R \cdot T}} \quad (2.22)$$

definiert. Constraint-basierte thermodynamische Methoden vereinfachen diesen Einfluss jedoch und betrachten nur, ob der Ablauf der Richtung einer Reaktion thermodynamisch möglich ist ($\gamma > 0$) oder nicht ($\gamma \leq 0$). Zentral für die hier vorgestellten thermodynamischen Methoden ist die Änderung der physiologischen Gibbs-Energie einer Reaktion (ab jetzt nur kurz „Gibbs-Energie“ genannt), die als

$$\Delta_r G' = \Delta_r G'^{\circ} + R \cdot T \cdot \sum_i (\eta_i \cdot \ln[P_i]) - R \cdot T \cdot \sum_j (\eta_j \cdot \ln[S_j]) \quad (2.23)$$

definiert ist und hier typischerweise die Einheit kJ/mol trägt. $\Delta_r G'^{\circ}$ steht hierbei für die Änderung der Standard-Gibbs-Energie einer Reaktion unter gegebenen physiologischen Bedingungen, im weiteren nur kurz „Standard-Gibbs-Energie“ genannt. Diese Energie bezieht sich auf eine Standard-Metabolit-Konzentration von 1 M sowie physiologisch passende Werte für relevante Daten wie die Wasseraktivität, den pH, die Magnesiumkonzentration (als pMg) und die Ionenstärke im Medium der Reaktion [45, 46]. R steht ferner für die Gaskonstante (standardmäßig, in den für diese Arbeit relevanten Einheiten, ca. $0,008314 \text{ kJ} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$), T für die Temperatur (in Kelvin, standardmäßig in den Berechnungen dieser Arbeit 298,15 K), $[P_i]$ bzw. η_i für die Konzentration bzw. Stöchiometrie aller Produkte und $[S_j]$ bzw. η_j für die Konzentration bzw. Stöchiometrie aller Substrate der Reaktion. Die Konzentrationen haben dabei in der vorliegenden Arbeit stets die Einheit M.

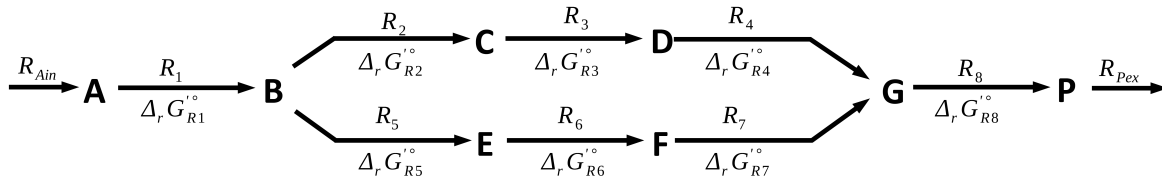
Ein wesentlicher Nutzen der Gibbs-Energie liegt in ihrer Verwendung als Indikator für die thermodynamische Möglichkeit einer Reaktion. Diese ist nur gegeben, wenn $\Delta_r G' < 0$. Andernfalls kann eine Reaktion nicht ablaufen. Ferner existiert die Definition der thermodynamischen Triebkraft einer Reaktion (Abb. 2.1A), die als

$$f_r = -\Delta_r G' \quad (2.24)$$

und somit als negierte Gibbs-Energie definiert ist. Folglich kann eine Reaktion nur ablaufen, wenn $f_r > 0$ kJ/mol. Das Konzept der thermodynamischen Triebkraft einer Reaktion kann nun als thermodynamische Triebkraft eines Stoffwechselweges verallgemeinert werden (Abb. 2.1B). Bei gegebenen fixen Metabolitkonzentrationen ist dies die minimale Triebkraft einer Reaktion, die im Stoffwechselweg auftritt. In [47] wurde dieses Konzept weiter verallgemeinert zur *maximal-minimalen* thermodynamischen Triebkraft eines Stoffwechselweges (Abb. 2.1C), kurz MDF (für Englisch *max-min driving force*). Für den MDF werden Bereiche von Metabolitkonzentrationen vorgegeben und daraus der MDF bestimmt. Nur wenn der MDF größer als 0 kJ/mol ist, ist der untersuchte Stoffwechselweg thermodynamisch möglich und kann ablaufen. Eine für diese Arbeit zentrale thermodynamische Parameter nutzende Methode der constraint-basierten Modellierung, OptMDFpathway [15], verallgemeinert die Definition des MDF nochmals. Anstatt nur den MDF eines vorgegebenen Pfades vorherzusagen, kann OptMDFpathway unter vorgegebenen

Flussnebenbedingungen den *netzwerkweiten* MDF (auch OptMDF genannt) zusammen mit einer zugehörigen Flussverteilung mit deren aktiven Reaktionen berechnen (Abb. 2.1D).

Beispielnetzwerk mit (vorgegebenen) Standard-Gibbs-Energien von Reaktionen:

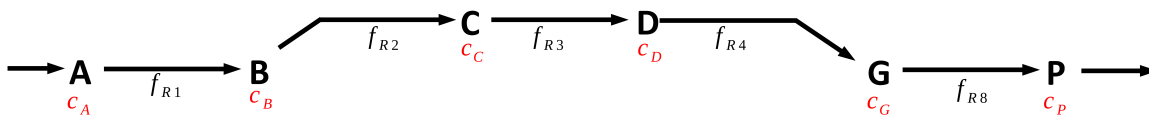


A) Triebkraft einer einzelnen Reaktion: $f_i = -\Delta_r G' = -(\Delta_r G^\circ + R \cdot T \cdot \sum (\eta_i \cdot \ln[P_i]) - R \cdot T \cdot \sum (\eta_j \cdot \ln[S_j]))$

Beispiel: $R_1(A \rightarrow B)$ mit **festen Konzentrationen:** $f_i = -\Delta_r G^\circ_{R1} - R \cdot T \cdot (-\ln(c_A) + \ln(c_B))$

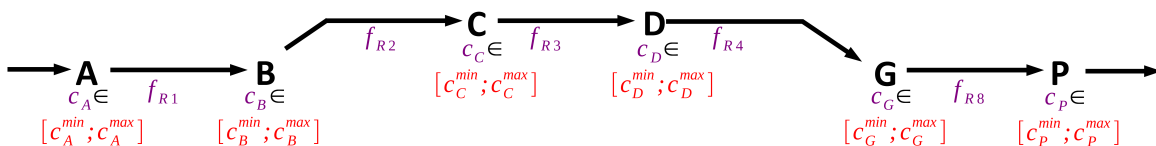
B) Triebkraft eines Weges: Minimale Triebkraft aller Reaktionen des Weges

$f_{\text{Weg}} = \min(f_{R1}, f_{R2}, f_{R3}, f_{R4}, f_{R8})$ mit **festen Konzentrationen**



C) Maximal-minimale Triebkraft (MDF) eines Weges: Maximal erreichbare Triebkraft des Weges

$MDF_{\text{Netzwerk}} = \max(f_{\text{Netzwerk}}) = \min(f_{R1}, f_{R2}, f_{R3}, f_{R4}, f_{R8})$ innerhalb gegebener **Konzentrationsbereiche**



D) MDF eines Netzwerkes: Finde Flussverteilung r welche den MDF aller aktiven Reaktionen maximiert

$MDF_{\text{Netzwerk}} = \max(MDF(r))$ innerhalb gegebener **Konzentrationsbereiche** und **Flussnebenbedingungen**

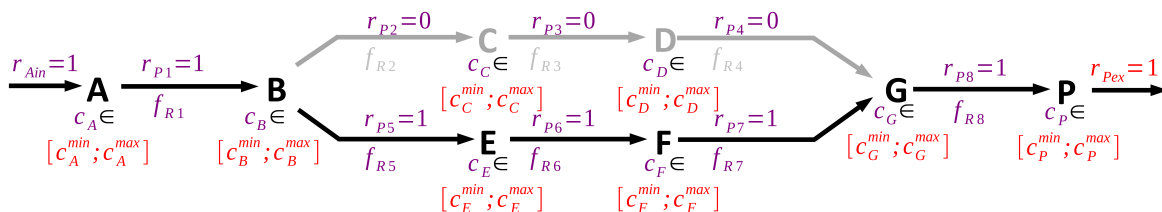


Abbildung 2.1: Überblick über, für diese Arbeit relevante, Definitionen im Zusammenhang mit thermodynamischen Triebkräften und dem daraus abgeleiteten MDF, illustriert durch ein kleines Beispielnetzwerk. (A) Triebkraft einer einzelnen Reaktion, (B) Triebkraft eines Weges, (C) MDF eines Weges und (D) MDF eines Netzwerkes. Rote Farben stehen für vorgegebene Nebenbedingungen, lila Farben für die Ergebnisse, die bei der Berechnung oder Optimierung des entsprechenden Wertes zurückgegeben werden. Für weitere Definitionen siehe den Abschnittstext.

Ähnlich wie die Triebkraft für eine einzelne Reaktion gibt der MDF für eine gesamte Menge aktiver Reaktionen unter anderem an, ob diese als Ganzes thermodynamisch möglich ($MDF > 0$ kJ/mol) oder nicht möglich ($MDF \leq 0$ kJ/mol) sind. Wann immer an jetzt in dieser Arbeit der MDF genannt wird, ist dieser *netzwerkweite* MDF gemeint.

Rechnerisch basiert OptMDFpathway auf den Nebenbedingungen der CBM (Gleichungen (2.9) bis (2.11)) und ergänzt diese um Nebenbedingungen, die auf der thermodynamischen Triebkraft basieren. Unter Ausnutzung des linearen Charakters von Gleichung (2.23) berechnet OptMDFpathway für jede Reaktion mit einer gegebenen Standard-Gibbs-Energie (für welche die Reaktion, um thermodynamische Plausibilität zu bewahren, massenbalanciert sein muss) die thermodynamische Triebkraft per

$$f_i = -\Delta_r G_i^\circ - R \cdot T \cdot N_{\cdot,i}^T \cdot \mathbf{x}. \quad (2.25)$$

$N_{\cdot,i}^T$ steht hierbei für die i -te Spalte der stöchiometrischen Matrix und \mathbf{x} für einen Vektor welcher die logarithmierte Konzentration jedes modellierten Metaboliten enthält. Um berechenbare Grenzen für die thermodynamische Triebkraft zu können, muss, für jede Konzentration in \mathbf{x} , eine Unter- und Obergrenze per

$$\ln(\mathbf{c}_{min}) \leq \mathbf{x} \leq \ln(\mathbf{c}_{max}) \quad (2.26)$$

gesetzt werden. \mathbf{c}_{min} und \mathbf{c}_{max} sind hierbei Vektoren, die für jeden Metaboliten die entsprechende Unter- und Obergrenze beinhalten. Zusätzlich ist es möglich, das Verhältnis zweier Metabolitkonzentrationen festzusetzen, indem

$$x_j - x_k = \ln(h) \quad (2.27)$$

gesetzt wird. j sowie k stehen hierbei für die Indizes der entsprechenden Metabolite, und h für das festzusetzende Konzentrationsverhältnis.

Das Ziel von OptMDFpathway ist die Ermittlung des schon beschriebenen netzwerkweiten MDF. Zum Nachvollziehen darüber, welche Reaktionen beim Erreichen dieses MDF des Netzwerks aktiv sind und welche nicht, wird für jede Reaktion mit einem $\Delta_r G^\circ$ eine Nebenbedingung der Form

$$r_i \leq z_i \cdot \beta_i \quad (2.28)$$

eingeführt. z_i steht in der Nebenbedingung für eine Binärvariable, welche also nur in der Menge

$$z_i \in \{0, 1\} \quad (2.29)$$

liegen kann. z_i ist dabei 0, wenn die damit assoziierte Reaktion i inaktiv ist, und 1, wenn i aktiv ist. Durch diese Binärvariablen stellt OptMDFpathway ein gemischt-ganzzahliges lineares Programm, kurz MILP (aus englisch *Mixed-Integer Linear Program*), dar. Pseudo-Reaktionen und Reaktionen ohne einen bekannten $\Delta_r G^\circ$ erhalten dabei keine Nebenbedingungen entsprechend der Gleichungen (2.28) und (2.29). Aufgrund der in dieser Arbeit genutzte Formulierung dieser Nebenbedingung müssen alle reversiblen Reaktionen in irreversible aufgeteilt werden, da ansonsten die Rückwärtsrichtung einer reversiblen Reaktion unabhängig von deren z_i ablaufen könnte.

Der MDF selbst wird durch eine kontinuierliche Variable B dargestellt. Dieser wird, entsprechend der Definition des MDF, per

$$B \leq f_i + M - z_i \cdot M \quad (2.30)$$

durch die minimale in aktiven Reaktionen auftretende Triebkraft beschränkt. M steht hierbei (im Sinne der verbreiteten „Big-M-Methode“ der linearen Programmierung) für einen sehr großen Zahlenwert, welcher zumindest größer als der höchstmögliche f_i sein sollte.

Um nun den MDF zu errechnen, wird eine Maximierung der Form

$$\begin{aligned} &\text{maximiere } B \\ &\quad \substack{r, x, z, B} \\ &u. d. N. (2.9) \text{ bis } (2.11), (2.25) \text{ bis } (2.30)' \end{aligned} \tag{2.31}$$

durchgeführt. Die zum MDF zugehörige Flussverteilung findet sich im Vektor r .

2.1.7 Minimal Cut Sets

Die Berechnung von Minimal Cut Sets (MCS) stellt eine Methode zur Vorhersage gezielter Eingriffe in einen Stoffwechsel dar. Eine wichtige Klasse von Eingriffen, die mit MCS vorhergesagt werden können, ist das schon beschriebene Ausschalten (Knock-Out) von Reaktionen und damit (biologisch interpretiert) implizit der Gene, welche die katalysierenden Enzyme der betroffenen Reaktionen kodieren.

Rechnerisch basiert die Bestimmung von MCS auf den Nebenbedingungen der CBM (Gleichungen (2.9) bis (2.11)). Zusätzlich kann eine *Target Region* (englisch für „Zielregion“) mit einzelnen Ungleichungen (bzw. Constraints) definiert werden. Jede dieser Ungleichungen soll durch die vorherzusagenden Knock-Outs der MCS *nicht* mehr möglich sein. Ein typisches Beispiel hierfür ist eine niedrige Produkt-zu-Substrat-Ausbeute. MCS selbst stellen nun eine jeweils minimale Anzahl an Interventionen dar, durch welche die Ziele der Target Region nicht mehr möglich sind, gleichzeitig aber ein in einer Desired Region beschriebene Verhalten weiterhin möglich bleiben muss. „Minimal“ ist hierbei so definiert, dass kein MCS eine Ober- oder Teilmenge eines anderen, ebenfalls die Ziele erreichenden, MCS sein darf.

Exakt definiert wird die Target Region im Rahmen von MCS-Berechnungen durch Nebenbedingungen der Form

$$\mathbf{T} \cdot \mathbf{r} \leq \mathbf{t}. \tag{2.32}$$

\mathbf{T} ist hierbei eine Matrix mit so vielen Spalten wie Reaktionen und Zeilen wie assoziierten Ungleichungen. \mathbf{t} ist ein Vektor, der für jedes Ungleichung der Target Region eine obere Schranke zuweist.

Neben der Target Region kann eine *Desired Region* (englisch etwa für „erwünschte Region“) definiert werden, durch welche die MCS zu eingeschränkten MCS (englisch *constrained MCS*) werden [48], die im Weiteren nur kurz MCS genannt werden. Die Desired Region beschreibt wiederum Ungleichungen welche möglich bleiben *müssen*. Ein typisches Beispiel ist hierbei eine minimale Wachstumsrate. Definiert wird die Desired Region analog zur Target Region durch

$$\mathbf{D} \cdot \mathbf{r} \leq \mathbf{d}. \tag{2.33}$$

\mathbf{D} ist hierbei eine Matrix mit so vielen Spalten wie Reaktionen und Zeilen wie assoziierten Ungleichungen. \mathbf{d} ist ein Vektor, der für jede Ungleichung der Desired Region eine obere Schranke zuweist.

Die Berechnung von MCS kann sehr aufwendig sein. Für größere Modelle metabolischer Netze ist die Berechnung aller MCS hierdurch nicht in angemessener Zeit möglich, sodass oft nur alle MCS bis zu einer bestimmten Anzahl an Interventionen berechnet werden können. Eine aktuell fortschrittliche Formulierung der MCS-Berechnung als MILP stellt die Arbeit in [27] dar. Diese erweitert die hier gegebene Definition von MCS um deutlich mehr, in dieser Arbeit nicht verwendeten, Optionen und beruht vor allem auf den Arbeiten in [49], [50] sowie [48].

2.1.8 Modellierung mikrobieller Gemeinschaften

Mittels metabolischer Modellierung lassen sich nicht nur einzelne Organismen, sondern ebenfalls Gemeinschaften aus Stämmen einer oder verschiedener Spezies darstellen [51–53]. Neben Methoden der kinetischen metabolischen Modellierung einer Gemeinschaft (z.B. [54, 55]), existieren auch auf der CBM beruhende Methodiken, u.a. [56–60].

In diesem Zusammenhang stellen OptDeg [61], SteadyCom [21] sowie RedCom [22] grundlegende stöchiometrisch konsistente Modellierungsmethoden für viele dieser Konzepte dar. Alle diese Methoden basieren wiederum auf dem Konzept des balancierten Gemeinschaftswachstums [62], wofür zunächst die stöchiometrische Matrix \mathbf{N} zur Gemeinschaftsmatrix \mathbf{N}^c (c für Englisch *community*) erweitert (Abb. 2.2). Jeder Stamm wird dabei in \mathbf{N}^c zu einem einzelnen Kompartiment des Gemeinschaftsmodells. Dadurch können z.B. Metabolite und Reaktionen, die in mehreren Stämmen vorkommen, für jeden Stamm getrennt betrachtet werden. Als zusätzliches Kompartiment wird das Austauschkompartiment (englisch *exchange compartment*) in das Gemeinschaftsmodell eingefügt (Abb. 2.2). Nur mit diesem haben die einzelnen Kompartimente der Stämme Austauschreaktionen. Außerdem verfügt nur das Austauschkompartiment über Austauschreaktionen mit der Umgebung. Somit entspricht \mathbf{N}^c der Form

$$\mathbf{N}^c = \begin{bmatrix} \mathbf{N}_1 & \dots & 0 & 0 \\ \vdots & \ddots & \vdots & 0 \\ 0 & \dots & \mathbf{N}_n & 0 \\ \mathbf{E}_{1,ex} & \dots & \mathbf{E}_{n,ex} & \mathbf{E}_c \end{bmatrix} \quad (2.34)$$

n ist hierbei die Anzahl der modellierten Stämme, \mathbf{N}_1 bis \mathbf{N}_n sind die ursprünglichen stöchiometrischen Matrizen eines jeden Stammes, $\mathbf{E}_{1,ex}$ bis $\mathbf{E}_{n,ex}$ die Austauschreaktionen der Stamm-Kompartimente und \mathbf{E}_c die Austauschreaktionen des Austauschkompartiments.

Die Nebenbedingung des Fließgleichgewichts der CBM (Gleichung (2.9)) ändert sich in einem Gemeinschaftsmodell somit zu

$$\mathbf{N}^c \cdot \mathbf{r}^c = \mathbf{0}, \quad (2.35)$$

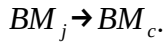
wobei \mathbf{r}^c den Flussvektor aller Reaktion des Gemeinschaftsmodells darstellt. Auch die weiteren Nebenbedingungen der CBM (Gleichungen (2.10) bis (2.11)) ändern sich analog zu

$$F_i \cdot \alpha_{ij} \leq r_{ij}^c \leq F_i \cdot \beta_{ij} \quad (2.36)$$

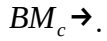
mit j als Index des jeweiligen Stammes der Gemeinschaft, und F_j als der Biomasse-Anteil des Stammes j an der gesamten Biomasse der Gemeinschaft, sowie

$$\mathbf{A}_i \cdot \mathbf{r}_i^c \leq F_i \cdot \mathbf{d}_i. \quad (2.37)$$

Eine spezielle Behandlung benötigt die Wachstums- bzw. Biomasse-Reaktionen der einzelnen Stämme. Zunächst wird hierfür die Biomasse-Reaktion jedes Stammes so umgewandelt, dass pro Wachstumsraten-Einheit ein Pseudo-Metabolit BM_j hergestellt wird. Jedes BM_j wiederum zum gemeinschaftlichen Pseudo-Metaboliten BM_c im Austauschkompartiment umgewandelt durch die neue Pseudo-Reaktion



BM_c wird schlussendlich in einer Austauschreaktion in die Umgebung überführt:



Entscheidend sind nun die Flussraten der BM_c produzierenden Pseudo-Reaktionen. Um stöchiometrisch konsistente Werte zu erreichen, müssen diese

$$r_{BM_j \rightarrow BM_c} = F_j \cdot \mu_c \quad (2.38)$$

folgen. μ_c ist die Wachstumsrate der gesamten Gemeinschaft. Für die F_j gilt dabei

$$\sum_{j=1}^n F_j = 1. \quad (2.39)$$

In Gleichung (2.38) wird angenommen, dass die Wachstumsrate für jeden Stamm der Gemeinschaft gleich ist. Dies wird angenommen, da nur eine solche Dynamik zu einer stabilen Gemeinschaft führt, in der kein Stamm einen anderen Stamm überwachsen kann. Bei Gleichung (2.38) ist dabei zu beachten, dass es sich um einen multiplikativen und damit nicht-linearen Term mit zwei Variablen handelt. Für constraint-basierte Modellierungsmethoden wird jedoch ein linearer Term benötigt. Daher muss entweder \mathbf{F} (unter Berücksichtigung von Gleichung (2.39)) oder μ_c auf einen konstanten Wert gesetzt werden.

Durch die Einführung der Biomasse-Anteile F_j müssen die in den einzelnen Stämmen auftauchenden Reaktionsflüsse entsprechend normalisiert werden. Dies erfolgt durch ein Umrechnen der Reaktionsgrenzen. Diese lauten nun

$$F_j \cdot \alpha_{ji} \leq r_i \leq F_j \cdot \beta_{ji}. \quad (2.40)$$

Mit konstanten F_j oder konstantem μ_c lassen sich nun beispielsweise Maximierungen und Minimierungen, wie bei der FBA, per

$$\underset{\mathbf{F}, \mathbf{r}}{\text{minimiere}} \mathbf{c}^T \cdot \mathbf{r}^c \quad (2.41)$$

u. d. N. (2.34) bis (2.40)

durchführen. Der Zielfunktionsvektor \mathbf{c}^T bezieht sich hierbei auf N^c (Gleichung (2.34)).

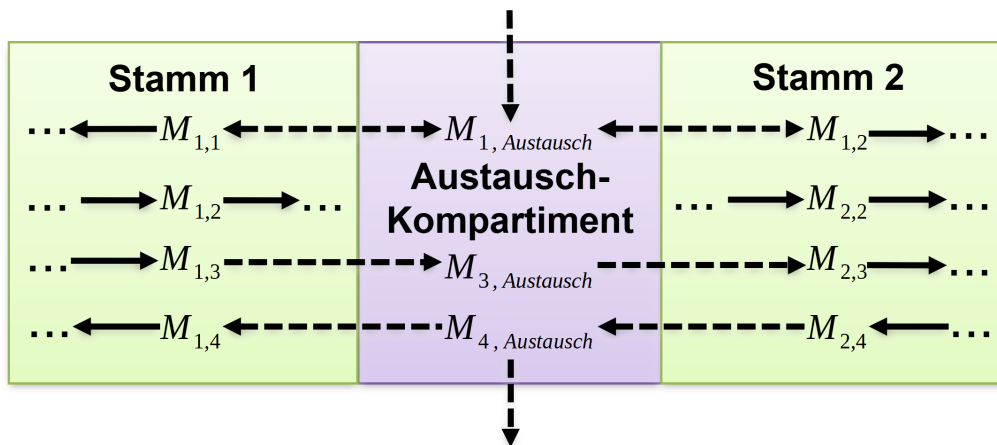


Abbildung 2.2: Struktur eines metabolischen Modells einer Gemeinschaft mehrerer Stämme einer oder mehrerer Spezies, zusammen mit möglichen Wegen der Austausch von Metaboliten. Gestrichelte Linien stehen für Austauschreaktionen und alle anderen für metabolische Umwandlungsreaktionen.

2.2 *Escherichia coli* und dessen Modellierung

2.2.1 Allgemeines und Stoffwechsel

Escherichia coli ist ein gramnegatives Bakterium aus der Klasse der γ -Proteobacteria. Seit dessen Entdeckung im Jahre 1885 [63] ist *E. coli* zum zentralen bakteriellen Modellorganismus der Molekular- und Systembiologie geworden. So ist sein Genom vollständig sequenziert [64] und ein Großteil gentechnischer Methoden wird an ihm entwickelt und getestet.

E. coli ist heterotroph und kann dutzende verschiedener Kohlenstoffsubstrate aufnehmen, darunter auch D-Glucose (im Folgenden nur Glucose genannt). Unter aeroben Bedingungen, d.h. bei Verfügbarkeit von Sauerstoff, wird Glucose u.a. in zentralen Teilen des Kohlenstoffstoffwechsels verarbeitet, der aus Komponenten wie der Glycolyse (mit der Gluconeogenese als ihrer nicht Glucose verwertenden reversen Form), dem Pentose-Phosphat-Weg und dem Citratzyklus besteht (für eine ausführlichere Gesamtbeschreibung siehe z.B. [65]).

Bei Sauerstoffverfügbarkeit (Aerobie) und hoher Glucoseverfügbarkeit im Wachstumsmedium neigt *E. coli* zusätzlich zum sogenannten „Überlaufstoffwechsel“ [66], auf Englisch *overflow metabolism*. Typisch für diesen Überlaufstoffwechsel ist die Produktion des organischen Endproduktes Acetat. Darüber hinaus nutzt *E. coli* unter Sauerstoffmangel (Anaerobie) Bedingungen zusätzliche Stoffwechselwege zur Sekretion organischer Endprodukte; Beispiele für solche Produkte sind Ethanol, Formiat und Succinat [67].

Wird *E. coli* unter geeigneten Bedingungen mindestens ein kohlenstoffhaltiges Substrat und alle essentiellen Nährstoffe zugesetzt, kann *E. coli* wachsen. Geschieht dies, nach einem Startzeitpunkt, ohne weitere Zugabe von Kohlenstoff, können die typischen bakteriellen Wachstumsphasen ablaufen [68]. Dabei kommt es nach einer langsamen „Latenzphase“ zu einer exponentiellen Wachstumsphase aufgrund der Vermehrung der Bakterien durch Zellteilung. Diese endet mit der stationären Phase, wenn kein Wachstumssubstrat mehr vorhanden ist oder auch zu viele toxische Endprodukte gebildet wurden.

2.2.2 Constraint-basierte Modelle von *E. coli*

Innerhalb der letzten Jahrzehnte wurden mehrere constraint-basierte metabolische Netzwerke von *E. coli* geschaffen. Alle für die vorliegende Arbeit relevanten Modelle leiten sich vom Modell *iJE660* ab [69], und lauten wie folgt:

- *iJO1366* [70]: Dieses genomskalige Modell enthält 2.583 Reaktionen, 1.805 Metabolite sowie 1.366 Gene.
- *EColiCore2* [71], kurz *ECC2*: Dieses Modell basiert auf *iJO1366* und wurde mithilfe eines speziellen Algorithmus zur Rationalisierung der Größe metabolischer Modelle [72] reduziert. Es enthält 499 Reaktionen sowie 486 (interne) Metabolite.
- *iML1515* [73]: Dieses genomskalige Modell stellt eine Aktualisierung und Erweiterung von *iJO1366* dar und besteht aus 2.712 Reaktionen, 1.877 Metaboliten sowie 1.516 Genen.

2.3 Wichtige informatische Projekte zur Stoffwechselmodellierung

2.3.1 Datenbanken und Formate metabolischer Modelle

Für viele Organismen können bereits existierende Daten zu metabolischen Netzen in darauf ausgelegten Datenbanken heruntergeladen werden. Eine dieser Datenbanken ist die Kyoto Encyclopedia of Genes and Genomes, kurz KEGG [74]. Diese beinhaltet für jeden beinhalteten Organismus Informationen zu seinem Genom, metabolischer Reaktionen sowie den dazugehörigen Metaboliten. Zu KEGG vergleichbare verbreitete Datenbanken, mit teils fehlenden und teils erweiterten Funktionen sowie Daten, sind MetaCyc [75] sowie MetaNetX [76]. Ferner existiert die Datenbank BiGG [77], welche komplett auf die constraint-basierte Modellierung ausgelegt ist und hierfür direkt in entsprechenden Programmpaketen nutzbare Modelle bereitstellt.

Jede dieser Datenbanken verfügt über ein eigenes System an Identifizieren für Gene, Metabolite sowie Reaktionen. So nutzt KEGG beispielsweise den Präfix „C“ für Metabolite, gefolgt von mehreren identifizierenden Ganzzahlen. BiGG hingegen nutzt einfacher zu lesende Abkürzungen aus den chemischen oder Kolloquial-Namen der Modellelemente. Bei Metaboliten in Modellen folgt bei BiGG zudem ein Unterstrich und die Abkürzung des Metabolit-Kompartiments.

Das am weitesten verbreitete Dateiformat für kinetische sowie constraint-basierte metabolische Modelle ist die XML-basierte Systems Biology Markup Language [78], kurz SBML. Die Referenzimplementierung zum Laden und Validieren von SBML-Dateien ist libsbml [79]. Diese wird für mehrere Programmiersprachen zur Verfügung gestellt und wird von nahezu allen SBML-nutzenden Programmpaketen der metabolischen Modellierung genutzt.

2.3.2 Software zur kinetischen Modellierung

Zwei weit verbreitete und für diese Arbeit relevante Softwareprojekte zur kinetischen Modellierung sind COPASI [80] sowie Tellurium [81]. COPASI bietet u.a. eine Vielzahl von Optimierungs- und Simulationsalgorithmen, die auch über eine graphische Benutzeroberfläche gesteuert werden können. Tellurium hingegen hat einen vergleichbaren Funktionsumfang wie COPASI, bietet aber keine graphische Oberfläche. Stattdessen können kinetische Modelle in einer eigenen

Beschreibungssprache namens Antimony formuliert und dann über das Python-Modul von Tellurium geladen und simuliert werden.

2.3.3 Software zur constraint-basierten Modellierung

Seit der ersten, in den 1980er-Jahren entwickelten, Software, die speziell für die constraint-basierte metabolische Modellierung entwickelt wurde [8], ist eine Vielzahl weiterer entsprechender Projekte entstanden. Projekte dieser Art können nach der verwendeten Programmiersprache kategorisiert werden:

- MATLAB: Allgemeine Projekte sind die „Constraint-Based Reconstruction and Analysis Toolbox“ [82], kurz COBRA Toolbox, und *CellNetAnalyzer* [11], kurz CNA. Beide Projekte beinhalten Methoden zur constraint-basierten Modellierung, darunter z.B. Methoden zur Berechnung von MCS. CNA zeichnet sich unter anderem dadurch aus, dass die Ergebnisse der Berechnungen direkt in interaktiven metabolischen Netzwerkkarten visualisiert werden können. Darüber hinaus existiert speziell für die Integration enzymatischer Nebenbedingungen und die Suche nach k_{cat} -Werten mit der GECKO-Methode die GECKO Toolbox [83].
- Python: Ein wichtiges Projekt in diesem Bereich ist COBRAPy [10], das von der COBRA Toolbox [82] inspiriert wurde. COBRAPy bietet Funktionen für alle grundlegenden Methoden der constraint-basierten Modellierung. Methoden zur Vorhersage genetischer Eingriffe bietet das Schwesterprojekt cameo [84]. Darüber hinaus existiert mit CNAPy [85] eine an CNA angelehnte graphische Oberfläche, die unter anderem auf COBRAPy aufbaut. Wie in CNA können auch in CNAPy Ergebnisse in interaktiven Karten dargestellt werden. Durch eine Einbindung des weiteren Python-Projektes StrainDesign [86] kann man in CNAPy zudem auf fortschrittliche Methoden des Metabolic Engineering zugreifen. Außerdem existiert mit Geckopy [87] eine Python-Alternative zur GECKO Toolbox. Des Weiteren existiert das Python-Paket equilibrator-api [45], mit dem $\Delta_r G^\circ$ von Stoffwechselreaktionen auf Basis der „Group Contribution“-Methode [46] berechnet werden können.

Diese Liste an Software-Projekten und Programmiersprachen ist nicht erschöpfend und beinhaltet nur einige der wichtigsten sowie die für diese Arbeiten relevanten Projekte und Sprachen.

3 Automatische Generierung constraint-basierter metabolischer Modelle mit effizient formulierten enzymatischen Nebenbedingungen

3.1 Motivation

Mit den schon dargestellten (siehe Abschnitt 2.1.5) constraint-basierten Methoden mit enzymatischen Nebenbedingungen MOMENT [12] sowie GECKO [13] konnten, wie in deren initialen Publikationen gezeigt, erfolgreiche Berechnungen und Vorhersagen durchgeführt werden. Dennoch besitzen beide Methoden Nachteile in deren Nutzung:

- MOMENT: Wie bereits erwähnt, besteht ein wesentlicher Nachteil von MOMENT darin, dass neu hinzugefügte enzymatische Nebenbedingungen nicht direkt in die stöchiometrische Matrix integriert werden. Dadurch können Modelle, die MOMENT verwenden, nicht mit gängigen Softwarepaketen zur constraint-basierten Modellierung verwendet werden.
- GECKO: GECKO löst zwar den genannten Nachteil von MOMENT, indem alle Nebenbedingungen in die stöchiometrische Matrix eingebaut werden. Gleichzeitig wird aber eine große Anzahl an Reaktionen sowie Metaboliten in die Matrix eingebaut. Dies kann die Komplexität von Berechnungen mit GECKO-erweiterten Modellen drastisch erhöhen. Insbesondere die Anwendung rechenintensiver Methoden, wie etwa die Berechnung von MCS, kann dadurch in angemessener Zeit unmöglich werden.

Aufgrund der genannten Nachteile ist es das Ziel der Arbeiten in diesem Kapitel, eine constraint-basierte Methode der Modellierung zu entwickeln welche a) in eine stöchiometrische Matrix einbaubar ist und b) gleichzeitig mit einer nur minimalen Vergrößerung der stöchiometrischen Matrix einhergeht. Die aus diesem Kapitel resultierende Methode heißt sMOMENT („short MOMENT“).

Zur praktischen Illustration der Fähigkeiten von sMOMENT ist es zudem das Ziel der Arbeit dieses Kapitels, ein um enzymatische Nebenbedingungen angereichertes genomskaliges Modell von *E. coli* zu erstellen. Dieses Modell soll sowohl unter aeroben als auch anaeroben Bedingungen ein realistisches Sekretionsverhalten vorhergesagt werden kann, einschließlich des Überlaufstoffwechsels. Dass dieser vom Gesamt-Enzympool abhängt, wird schon aufgrund von Messungen *in vivo* sowie grundlegender mathematischer Überlegungen angenommen [66]. Das aus dieser Arbeit resultierende Modell wird *iJO1366** genannt. Ein weiteres Ziel dieser Arbeit ist zudem eine programmatische Referenzimplementierung von sMOMENT, die mit dem Paket AutoPACMEN realisiert wurde.

3.2 Die sMOMENT-Methode

Wie MOMENT und GECKO basiert auch sMOMENT auf den grundlegenden Nebenbedingungen der CBM (Gleichungen (2.9) bis (2.11)). Im Hinblick auf die bereits genannten Ziele wird sMOMENT jedoch auf einem anderen Weg als bei MOMENT bzw. GECKO hergeleitet. Hintergrund dieser modifizierten Herleitung ist zunächst, dass vereinfachend angenommen wird,

dass jedes modellierte Enzym aus dem Gesamt-Enzympool entnommen wird. Darüber hinaus wird Gleichung (2.17) zu

$$\frac{r_i}{k_{cat,i}} \leq [E_i] \quad (3.1)$$

umgeformt. Zudem wird anstatt von Gleichung (2.18), wo von Enzymkonzentrationen ausgegangen wird, für jedes Enzym i die Nebenbedingung des Gesamt-Enzympools Ω (in $\text{g} \cdot \text{gDW}^{-1}$) direkt von der Gesamt-Enzympool-Entnahme g_i (in $\text{mmol} \cdot \text{gDW}^{-1}$) sowie vom Molekulargewicht eines Enzyms, W_i (in $\text{g} \cdot \text{mmol}^{-1}$), ausgegangen:

$$\sum_i^n g_i \cdot W_i \leq \Omega \quad (3.2)$$

g_i kann somit auch als Enzymkonzentration $[E_i]$ betrachtet werden, und wird hier nur semantisch voneinander unterschieden um das Nutzen des Gesamt-Enzympools zu verdeutlichen. Gleichungen (3.1) sowie (3.2) lassen sich nun als

$$\sum_i^n r_i \cdot \frac{W_i}{k_{cat,i}} \leq \sum_i^n g_i \cdot W_i \leq \Omega \quad (3.3)$$

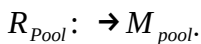
zusammenfassen, oder noch kürzer als

$$\sum_i^n r_i \cdot \frac{W_i}{k_{cat,i}} \leq \Omega. \quad (3.4)$$

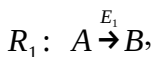
Unter der Hinzunahme einer Pseudo-Reaktion im Modell, welche global Proteinmasse an jede einzelne Reaktion mit einem Fluss r_{gesamt} liefert, lässt sich Gleichung (3.4) auch wie folgt darstellen:

$$-\sum_i^n r_i \cdot \frac{W_i}{k_{cat,i}} + r_{gesamt} = 0; r_{gesamt} \leq \Omega \quad (3.5)$$

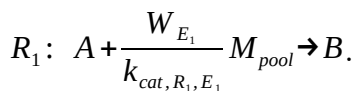
Daraus ergeben sich einige minimale Erweiterungen der stöchiometrischen Matrix, um alle modellierten enzymatischen Nebenbedingungen zu integrieren. Zum einen wird der Gesamt-Enzympool, wie bei GECKO, als Pseudo-Metabolit durch eine Pool-Pseudoreaktion bereitgestellt:



Im Gegensatz zu GECKO kann jedoch bei Anwendung von Gleichung (3.5) auf die Konversionsreaktionen vom Pool zu den einzelnen Enzymen verzichtet werden. Stattdessen kann jede Reaktion direkt auf den gesamten Enzympool zugreifen. In dem bereits verwendeten Beispiel einer einfachen Reaktion



welche durch ein einzelnes Enzym katalysiert wird, wird diese in sMOMENT wie folgt umgeschrieben:



Damit diese einfache Zuweisung ebenfalls für komplexere Enzym-Reaktions-Verhältnisse gilt, wird formal zunächst die Menge aller im Modell auftauchenden q Enzyme

$$E = \{E_1, \dots, E_q\} \quad (3.6)$$

ausgegangen. Aus dieser Menge lässt sich nun für jede Reaktion i eine eigene Teilmenge

$$E(i) = \{E_{i,1}, E_{i,2}, \dots\} \quad (3.7)$$

bilden welche alle Enzyme beinhaltet, die in der Lage sind die Reaktion i zu katalysieren. Für jede Reaktion des Modells wird nun, als zentraler Schritt von sMOMENT, das minimale Verhältnis aus Molekulargewicht des Enzyms zum dazugehörigen k_{cat} über

$$\xi_i = \frac{W_i}{k_{cat,i}} = \min\left(\left\{\frac{W_{E_{i,1}}}{k_{cat,E_{i,1}}}, \frac{W_{E_{i,2}}}{k_{cat,E_{i,2}}}, \dots\right\}\right) \quad (3.8)$$

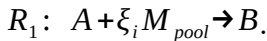
berechnet, wobei

$$E_{i,1}, E_{i,2}, \dots \in E(i) \quad (3.9)$$

gilt. Im Falle, dass ein Enzym θ tatsächlich einen Enzymkomplex darstellt, gilt überdies

$$W_{E_\theta} = \sum_{\iota \in \Psi_\theta} \eta_{E_\iota, \theta} \cdot W_{E_\iota, \theta} \quad (3.10)$$

gilt, wobei $\eta_{E_\iota, \theta}$ für die interne Stöchiometrie des Unterenzyms ι im Enzymkomplex und Ψ_θ für die Menge aller Indizes der Enzyme des Komplexes steht. Das in Gleichung (3.8) gegebene minimale Verhältnis wird nun als Koeffizient für den Enzympool eingebaut, beispielsweise in der schon genutzten einfachen Beispielreaktion als



Der Grund für die Wahl dieses minimalen Verhältnisses liegt in der Natur der Optimierung mit enzymatischen Nebenbedingungen. Wenn z.B. in einem Modell mit enzymatischen Nebenbedingungen ein Fluss, etwa die Wachstumsrate, maximiert wird, so wird die optimale Lösung immer eine maximale relevante Ausbeute aus dem Gesamt-Enzympool enthalten. Wenn mehrere Enzyme dieselbe Reaktion katalysieren, wird somit immer das Enzym mit dem niedrigsten Verhältnis von Molekulargewicht zu k_{cat} ausgewählt. Nur diese Reaktion wird in ein sMOMENT-Modell aufgenommen, ohne dass dies folglich einen Einfluss auf das Ergebnis einer typischen Optimierung hat.

Zusammengefasst werden somit in sMOMENT, ohne eingebaute gemessene Proteinkonzentrationen, lediglich *ein* Pseudo-Metabolit und nur *eine* Pseudo-Reaktion an die ursprüngliche stöchiometrische Matrix angefügt (Abb. 3.1). Dies steht im starken Kontrast zu GECKO, wo für jedes modellierte Enzym als eigenständiger Metabolit eingefügt und hierfür zumindest eine jeweilige Umwandlungsreaktion hinzugefügt wird. Trotz dieser starken Vereinfachung hat sMOMENT, wie beschrieben, für die Anwendung typischer Optimierungsprobleme im Rahmen der CBM die gleiche Aussagekraft wie MOMENT sowie GECKO.

$$\begin{array}{c}
M_1 \\
\vdots \\
M_m \\
\hline
M_{pool}
\end{array}
\left[\begin{array}{ccc|c}
R_1 & \dots & R_n & R_{pool} \\
\eta_{1,1} & \dots & \eta_{1,n} & 0 \\
\vdots & \ddots & \vdots & 0 \\
\eta_{m,1} & \dots & \eta_{m,n} & 0 \\
\hline
-\frac{W_1}{k_{cat,1}} & \dots & -\frac{W_n}{k_{cat,n}} & 1
\end{array} \right]$$

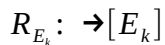
$$r_{R_{pool}} \leq \Omega$$

Abbildung 3.1: Erweiterung der stöchiometrischen Matrix N durch sMOMENT ohne Nutzung gemessener Proteinkonzentrationen. M_{pool} ist der Pseudo-Metabolit des Gesamt-Enzympools und R_{pool} die Pseudo-Reaktion die diesen erzeugt. R_i steht für Reaktion i , M_j für Metabolit j , n ist die Anzahl Reaktionen, m die Anzahl Metabolite und η die jeweilige Stöchiometrie eines Metaboliten.

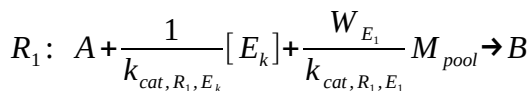
Mehr als eine zusätzliche Pseudo-Reaktion und ein zusätzlicher Pseudo-Metabolit sind bei sMOMENT nur bei der Integration bekannter Enzymkonzentrationen notwendig. Hierbei werden auf Gleichung (3.1) beruhende, und die Enzymkonzentration direkt als Schranke beinhaltende, Nebenbedingungen folgender Form eingebaut:

$$\sum_{i \in R(E_k)} \frac{r_i}{k_{cat,i}} \leq [E_k] \quad (3.11)$$

Hierbei steht E_k für das jeweilige gemessene Enzym und $R(E_k)$ für alle Reaktionen die durch das Enzym katalysiert werden. $[E_k]$ ist nun ein eigener Pseudo-Metabolit der über eine Pseudo-Reaktion der Form



mit einer vorgegebenen unteren Flussgrenze von 0 und einer oberen Grenze entsprechend der Enzymkonzentration in $\text{mmol} \cdot \text{gDW}^{-1}$. Die schon gegebene sMOMENT-Beispielreaktion würde sich im Falle einer gemessenen Enzymkonzentration (siehe nächsten Abschnitt) gemäß der sMOMENT-Referenzimplementierung zu



ändern. Hierbei ist zu bedenken, dass auch gleichzeitig eine Entnahme aus dem Gesamt-Enzympool stattfindet, weswegen dieser, anders als etwa in GECKO, auch die Menge der gemessenen Enzyme beinhalten muss.

3.3 AutoPACMEN

Die im Rahmen dieses Kapitels erstellte Referenzimplementierung von sMOMENT heißt „Automatic integration of Protein Allocation Constraints in Metabolic Networks“, kurz AutoPACMEN. Es besteht aus zwei Hauptkomponenten mit dem erweiterten Ziel, nicht nur das Einbauen von sMOMENT-Nebenbedingungen, sondern auch das Sammeln und Kalibrieren aller dafür notwendiger enzymatischer Daten zu automatisieren. Letzteres geschieht zudem mit einem im Vergleich zur GECKO Toolbox verfeinerten Algorithmus.

Die erste Hauptkomponente von AutoPACMEN ist der Modellgenerator (Abb. 3.2). Er ist in Python geschrieben und basiert auf COBRAPy [10]. Obligatorische Eingaben sind die SBML-Datei eines constraint-basierten Stoffwechselmodells und ein taxonomisch gültiger Organismusname. Die SBML-Datei muss dabei BiGG-Identifizierer für Metaboliten verwenden, EC-Nummer-Annotationen für Reaktionen und Enzym-Reaktions-Regeln mit UniProt-kompatiblen Enzymnamen (oder assoziierten Gennamen) enthalten. Schon fertiggestellte Modelle im SBML-Format und mit dieser Form an Zusatzdaten findet man bspw. in der BiGG-Datenbank [77]. Optionale Eingaben umfassen, jeweils im JSON-Format, eine manuell erstellte k_{cat} -Datenbank und eine Liste der gemessenen Enzymkonzentrationen. Auf der Grundlage dieser Eingaben wählt der Modellgenerator k_{cat} -Werte für alle Modellreaktionen aus, für die es eine auslesbare EC-Nummern-Annotation gibt. Dies geschieht, im Rahmen des in Algorithmus 3.1 beschriebenen programmtechnischen Ablaufs des Modellgenerators, unter Verwendung der Datenbanken BRENDA [88] und SABIO-RK [89] sowie gegebenenfalls der optionalen benutzerdefinierten Datenbank. Der in Algorithmus 3.1 beschriebene Ansatz integriert zudem taxonomische Informationen über die Datenbank NCBI Taxonomy [90], um Werte von taxonomisch möglichst nahen Spezies zu erhalten. Darüber hinaus verwendet AutoPACMEN UniProt [91], um die Molekulargewichte der Enzyme innerhalb des Modells abzuleiten. Zusätzlich generiert AutoPACMEN verschiedene editierbare Tabellendateien, die es dem Benutzer ermöglichen, die Größe des gesamten Enzympools und die internen Stöchiometrien der Enzymkomplexe bequem einzustellen. Mit all diesen Informationen erstellt der Modellgenerator ein constraint-basiertes Stoffwechselmodell im SBML-Format, das durch sMOMENT-Nebenbedingungen erweitert wird. Für die vergleichende Analyse enthält AutoPACMEN eine zu GECKO analoge Methode zur Modellgenerierung, die die gleiche Anzahl an addierten Reaktionen und Metaboliten liefert wie GECKO.

Die zweite Hauptkomponente von AutoPACMEN ist der Modellkalibrator, der eine Anpassung des Gesamt-Enzympools und der k_{cat} -Werte durchführen kann. Diese Komponente wurde teils in Python und teils in MATLAB geschrieben (Abb. 3.2). Der Python-Teil nutzt, wie der Modellgenerator, COBRAPy, während der MATLAB-Teil *CellNetAnalyzer* [11] verwendet. Der Modellkalibrator nimmt mehrere Eingaben entgegen: Als Eingabe obligatorisch ist dabei ein sMOMENT-Modell, das vom Modellgenerator im SBML-Format erzeugt wurde. Eine weitere obligatorische Eingabe besteht aus einer vom Benutzer erstellten Liste von Optimierungsszenarien im JSON-Format. Jedes Szenario enthält den Identifizierer einer zu optimierenden Reaktion, ein erwartetes Optimierungsziel in Form eines Flusses, das durch die Kalibrierung erreicht werden können soll, veränderbare Flussgrenzen der jeweiligen Reaktion und einen Gewichtungswert, der während der Kalibrierung verwendet wird. Diese Eingaben dienen den beiden grundlegenden

Funktionen des Modellkalibrators, der Kalibrierung des Gesamt-Enzym-pools und der Kalibrierung der k_{cat} -Werte. Für die Kalibrierung der k_{cat} -Werte ist eine zusätzliche Eingabe in Form einer JSON-Datei erforderlich, die die Identifizierer der Reaktionen enthält, deren k_{cat} -Werte kalibriert werden sollen. Die Verwendung einer solchen vorausgewählten Liste kann den Kalibrierungsprozess erheblich beschleunigen und kann über eine k_{cat} -Sensitivitätsanalyse mit dem Python-Skript des Modellkalibrators automatisch erstellt werden. Dieses Skript hilft dabei, k_{cat} -Werte zu identifizieren, die sich ausschließlich auf den Optimierungswert in einem einzelnen der angegebenen Optimierungsszenarien auswirken. In Hinblick auf alle Eingaben führt der Modellkalibrator eine nichtlineare Optimierung, unter Nutzung der `fmincon`-Funktion von MATLAB, sowohl für die Enzym-pool- als auch für die k_{cat} -Kalibrierung anhand der folgenden Optimierung durch:

$$\text{minimiere } \sum_j \sum_{r_{ij}^{\zeta}} w_{ij} \cdot \max\left(\frac{r_{ij}^{\zeta}}{r_{ij}^{\zeta}}, \frac{r_{ij}^{\zeta}}{r_{ij}^{\zeta}}\right)$$

Hierbei steht r_{ij}^{ζ} für den, im Rahmen der Optimierung durchgeführten FBA, ermittelten Fluss der Reaktion i im Optimierungsszenario j , r_{ij}^{ζ} für den entsprechenden im Szenario gesetzten zu erwartenden Reaktionsfluss und w_{ij} für die Gewichtung des Szenarios. Durch die konkrete Formulierung der Optimierung wird sichergestellt, dass die Summe der *relativen* Abweichungen vom originalen Enzym-pool oder k_{cat} minimiert wird, sodass u.a. höhere Zielwerte keinen entsprechenden höheren Einfluss auf die Optimierung haben.

AutoPACMEN und alle hier dargelegten Berechnungen sowie das hier beschriebene Modell *iJO1366** sind unter dem folgenden GitHub-Repository veröffentlicht:

<https://github.com/klamt-lab/autopacmen>

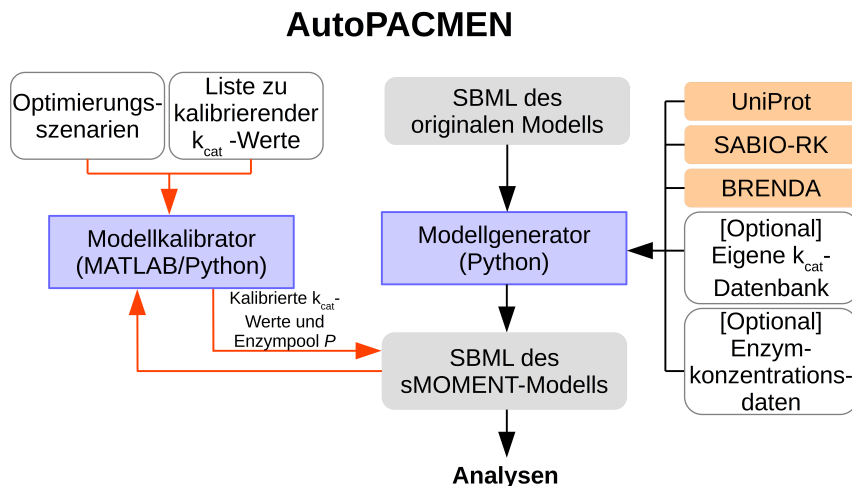


Abbildung 3.2: Übersicht über den Aufbau und den Arbeitsablauf mit AutoPACMEN. Die roten Pfeile stellen die Verwendung des optionalen Modellkalibrators dar. Die grauen Kästen stehen für Eingabe- und Ausgabedateien, die blauen für die zentralen Komponenten von AutoPACMEN, die orangenen für externe Datenbanken, aus denen Daten gelesen werden, und die weißen für optionale Daten, die vom Benutzer zur Verfügung gestellt werden.

Für jede Reaktion R des Modells, führe Folgendes aus:

Schritt 1:

- I. Sammle alle EC-Nummern der Reaktion aus der Modellbeschreibung als Liste **A**.
- II. Lösche aus **A** alle EC-Nummern die Platzhalter-Symbole wie „*“ oder „-“ enthalten. (Dies wird getan, um das Einführen von k_{cat} -Werten aus unterschiedlichen Reaktionen zu verhindern.)
- III. Sammle die BiGG-Identifizierer aller Substrate der Reaktion als Liste **B**.
- IV. Wenn **A** nach allen bisherigen Schritten leer ist: Gehe zu Schritt 2A. Ansonsten gehe zu Schritt 2B.

Schritt 2A (Für die aktuelle Reaktion R):

- I. Sammle die Liste aller Enzyme bzw. Enzymkomplexe der Reaktion anhand ihrer Gen-Reaktions-Regel als Liste **C**.
- II. Erstelle eine leere Liste **D**.
- III. Für jedes Enzym aus **C**: Überprüfe, ob dieses in der benutzereigenen k_{cat} -Datenbank einen k_{cat} -Eintrag für das Enzym existiert. Wenn kein k_{cat} -Eintrag gefunden werden konnte, fahre mit Schritt 3 fort. Andernfalls sammle alle k_{cat} -Einträge, wähle den maximalen k_{cat} aus den Einträgen aus und füge diesen der Liste **D** an.
- IV. Wähle den maximalen Wert aus **D** als k_{cat} für R aus.
- V. Fahre mit der nächsten Reaktion und Schritt 1 fort.

Schritt 2B (Für die aktuelle Reaktion R):

- I. Finde alle k_{cat} -Einträge in SABIO-RK sowie BRENDA für alle Einträge aus **A** mit einer EC-Nummer von R und speichere diese als **E**. Diese Einträge enthalten neben dem k_{cat} auch den Organismus und das Substrat in dem der Wert gemessen wurde.
- II. Falls so auslesbar und möglich, ändere für jeden Eintrag in **E** ob den Namen vom Substrat in dessen BiGG-Metabolit-Identifizierer. Wenn dies für einen Eintrag möglich war, erhalte nur die Einträge mit dem Substrat (bei der Vorwärtsrichtung) oder dem Produkt (bei der Rückwärtsrichtung), und speichere dies als **F**. Wenn dies nicht möglich war, kombiniere alle Einträge in einen Meta-Eintrag ohne BiGG-Identifizierer als **F**.
- III. Ermittle, mithilfe von NCBI TAXONOMY, für jeden Eintrag in **F** die taxonomische Distanz des Organismus im Eintrag zum modellierten Organismus, und speichere diese Distanz im Eintrag ab.
- IV. Sammle nun alle k_{cat} -Einträge aus **F** mit der geringsten ermittelten taxonomischen Distanz, und speichere dies als **G**.
- V. Wenn **G** kleiner als 10 Einträge ist, fahre mit der nächsthöheren taxonomischen Distanz fort und füge die entsprechenden Einträge zu **G** zu. Wiederhole diesen Schritt, solange **G** weiterhin kleiner als 10 Einträge ist oder die höchste taxonomische Distanz erreicht ist.
- VI. Führe Schritt 2A aus ohne Unterschnitt V. Wenn **D** Einträge enthält, füge sie zu **G** hinzu.
- VII. Berechne den Median aller Einträge in **G** und wähle diesen als k_{cat} -Wert von R aus. Sollte die Menge **G** leer sein, gehe zu Schritt 3. Ansonsten fahre mit Schritt 1 und der nächsten Reaktion fort.

Schritt 3: (nur falls kein k_{cat} der Reaktion R gefunden werden konnte)

- I. Setze für R , nachdem alle Reaktionen durchgegangen wurden, den Median aller gesetzten k_{cat} -Werte.

Algorithmus 3.1: Auswahlmechanismus der reaktionsabhängigen k_{cat} -Werte für ein metabolisches Modell durch AutoPACMEN.

3.4 Anwendung von AutoPACMEN an einem *E. coli*-Modell

Als beispielhafte Anwendung wurde sMOMENT mithilfe von AutoPACMEN auf das genomskalige *E. coli*-Modell *iJO1366* [70] angewendet. Das resultierende Modell wurde als *iJO1366** bezeichnet. Mit dem Ziel, eine biologische Einschränkung der Wachstumsrate allein durch enzymatische Nebenbedingungen zu erreichen und deren Auswirkung besser beurteilen zu können, sollte das Wachstum von *iJO1366** zudem ausschließlich durch enzymatische Nebenbedingungen begrenzt werden.

3.4.1 Erstellung & Kalibrierung des Modells

Genutzte Softwareversionen

Alle im Folgenden dargestellten Berechnungen erfolgten mit Python 3.7, MATLAB 2017a, IBM CPLEX 12.8, *CellNetAnalyzer* 2019.1 sowie COBRAPy 0.15.

Datenquellen

Das originale *iJO1366*-Modell wurde aus der BiGG-Datenbank [77] als SBML-Datei bezogen.

Neben den in AutoPACMEN standardmäßigen k_{cat} -Quellen BRENDA sowie SABIO-RK wurde als zusätzliche Quelle für direkt in *E. coli* gemessene k_{cat} -Werte [92] verwendet. Diese Quelle beinhaltet k_{app} , also „scheinbare“ (englisch *apparent*) k_{cat} . Solche „scheinbaren“ Werte wurden direkt *in vivo* aus gemessenen Flüssen und Proteinkonzentrationen unter mittlerem Wachstum bei Glucose sowie aeroben Bedingungen errechnet, im Kontrast zu den unter kontrollierteren Bedingungen und oftmals dafür *in vitro* gemessenen k_{cat} .

Zwei verschiedene Datenquellen wurden für die Wachstumsraten benutzt. Für aerobe sowie anaerobe Werte mit Glucose diente [93] als Quelle, welche zudem als Quelle für die Sekretion typischer Fermentationsprodukte unter aeroben sowie anaeroben Bedingungen diente. Weitere aerobe Wachstumsraten für 23 andere Kohlenstoffquellen wurden aus [12] entnommen.

Erstellung eines sMOMENT-Modells

Ausgehend vom originalen *iJO1366* umfasste die erste Phase der Erstellung von *iJO1366** den Einbau der sMOMENT-Nebenbedingungen unter Verwendung von AutoPACMEN und den angegebenen k_{cat} - und k_{app} -Quellen. Zur Vereinfachung wurde für alle modellierten Enzymkomplexe eine interne Stöchiometrie von 1 angenommen. Um sicherzustellen, dass die Glucose-Aufnahme ausschließlich durch enzymatische Nebenbedingungen eingeschränkt wird, wurde außerdem die maximale Glucose-Aufnahmerate freigesetzt. Für die taxonomisch orientierte k_{cat} -Suche mittels AutoPACMEN wurde der Begriff „*Escherichia coli*“ verwendet.

Das aus diesem Schritt resultierende Modell heißt *iJO1366*A* und besitzt, im Vergleich zum originalen *iJO1366*, 595 Reaktionen und einen Metaboliten mehr (Tab. 3.1). Alle neuen Reaktionen, bis auf diejenige die den Gesamt-Enzympool liefert, resultierten daraus, dass für sMOMENT alle reversiblen Reaktionen mit gefundenen Enzymdaten irreversibel aufgeteilt werden mussten. Der eine neu hinzugefügte Metabolit ist der Pseudo-Metabolit des Gesamt-Enzympools.

Tabelle 3.1: Modellgrößen des originalen *iJO1366* im Vergleich zum *sMOMENT*-erweiterten *iJO1366*A*.

	<i>iJO1366</i>	<i>iJO1366*A</i>
Anzahl an Reaktionen	2.583	3.178
Anzahl an Metaboliten	1.805	1.806

Enzympool-Kalibrierung

Der nächste Schritt umfasste die Durchführung einer Kalibrierung des Gesamt-Enzympools für *iJO1366A*, was im Modell *iJO1366B* resultierte. Zunächst wurde hierfür ein „Standard-Austauschscenario“ festgelegt, das für alle nachfolgenden Berechnungen und Modelle verwendet wurde. In diesem Szenario war nur das jeweils gesetzte Kohlenstoff-Substrat als solches Substrat erlaubt. Zusätzlich wurden nur die typischen Sekretionsmetabolite Acetat, Ethanol, Formiat, Succinat, Lactat sowie CO₂ als ausscheidbare kohlenstoffhaltige Stoffe erlaubt. Alle anderen standardmäßig in *iJO1366* gesetzten Austausche von Metaboliten mit der Umgebung verblieben unverändert. Als Datengrundlage für die Kalibrierung des Gesamt-Enzympools dienten zwei maximale Wachstumsraten von *E. coli* aus [93]. Diese wurden unter aeroben sowie anaeroben Bedingungen mit Glucose als Kohlenstoffquelle gemessen (Tab. 3.2).

Aufgrund der kleinen Datenmenge erfolgte die Kalibrierung manuell und resultierte in einem optimalen Enzympool von 0,095 g·gDW⁻¹ für *iJO1366*B*. Die daraus resultierende Differenz zwischen der optimalen Wachstumsrate mit dem Modell und den Mittelwerten aus [93] beträgt aerob 0 h⁻¹ und anaerob 0,05 h⁻¹ (Tab. 3.2).

Tabelle 3.2: Vergleich der von Glucose abhängigen Wachstumsraten von *iJO1366*B* mit den *in vivo* gemessenen Werten in [93].

Bedingung	Enzympool-optimiertes <i>iJO1366*B</i> Maximale Wachstumsrate [h ⁻¹]	Monk et al., 2016 Gemessene Wachstumsrate mit Standardabweichung [h ⁻¹]
Glucose, aerob	0,73	0,73 ± 0,01
Glucose, anaerob	0,51	0,46 ± 0,02

Manuelle k_{cat} -Anpassungen

FBA- und FVA-basierte Analysen mit *iJO1366*B* unter anaeroben Bedingungen zeigten eine deutliche Abweichung in der vorhergesagten Sekretion typischer Fermentationsprodukte im Vergleich zu *in-vivo*-Daten aus [93] bei einer Glucose-Aufnahmerate von 16,69 mmol·gDW⁻¹·h⁻¹. Um die in den nachfolgenden Abschnitten beschriebenen k_{cat} -Anpassungen zu erleichtern, wurde eine manuelle Anpassung der k_{cat} -Werte für zentrale Fermentationsreaktionen durchgeführt, die unplausibel hohe oder niedrige Aktivitäten zeigten. Konkret wurden hierfür 4 manuelle Anpassungen vorgenommen:

- Der k_{cat} der Reaktion PFL (Pyruvat-Formiat-Lyase) wurde um das 10-fache erhöht.

- Der k_{cat} der Rückwärtsrichtung der Reaktion ACALD (Acetylierende Acetaldehyd-Dehydrogenase) wurde um das 5-fache erhöht.
- Der k_{cat} der Rückwärtsrichtung der Reaktion ALCD2x (Ethanol-Dehydrogenase) wurde um das 10-fache erniedrigt.
- Der k_{cat} der Rückwärtsrichtung der Reaktion LDH_D (D-Lactat-Dehydrogenase) wurde um das 10-fache erniedrigt.

Als Resultat der manuellen Anpassungen ergab sich das Modell *iJO1366*C*. Die Vorhersage des Sekretionsverhalten dieses Modell hat sich im Vergleich zu den Resultaten mit *iJO1366*B* erheblich verbessert (Tab. 3.3).

Table 3.3: Anaerobe Sekretionsraten von *iJO1366*B* sowie *iJO1366*C* im Vergleich zu den in vivo gemessenen Werten von [93]. FVA über die Sekretionsreaktionen zeigten, dass lediglich eine vernachlässigbar kleine Variabilität in diesen Raten besteht.

Produkt	Anaerobe Sekretionsraten [mmol · gDW ⁻¹ · h ⁻¹] bei einer Glucose-Aufnahmerate von 16,69 mmol · gDW ⁻¹ · h ⁻¹		
	<i>Monk et al., 2016 (in vivo)</i>	<i>sMOMENT-Modell iJO1366*B</i>	<i>Optimiertes sMOMENT-Modell iJO1366*C</i>
Ethanol	11,22 ± 0,6	2,23	13,3
Formiat	22,17 ± 1,69	6,14	28,6
Acetat	11,71 ± 1,14	2,42	13,5
Lactat	0 ± 0	23,2	0
Succinat	1,86 ± 0,4	0,12	0,141

Automatisierte k_{cat} -Kalibrierung

Trotz der manuellen k_{cat} -Änderungen, die in *iJO1366*C* resultierten, bestand weiterhin eine starke Diskrepanz zwischen den gemessenen Wachstumsraten unter 23 verschiedenen Kohlenstoffquellen aus [12] und den vorhergesagten Wachstumsraten. Daher wurde mittels des AutoPACMAN-Modellkalibrierers eine automatisierte k_{cat} -Kalibrierung durchgeführt.

Als Szenarien, nach denen kalibriert wird, dienten die genannten 23 Wachstumsraten aus [12] sowie die aerobe und anaerobe Wachstumsrate aus [93]. Die beiden letzten Werte erhielten, als einzige Werte eines Experiments bei dem sowohl aerobe als auch anaerobe Bedingungen betrachtet wurden, eine Kalibrierungs-Gewichtung von 3. Alle anderen Szenarien erhielten eine Gewichtung von 1.

In einem ersten Schritt wurden die k_{cat} des Modells ermittelt, deren Änderung einen Einfluss auf lediglich eines der insgesamt 25 Wachstums-Szenarien hat. Durch diese Vorauswahl sollte die nachfolgende automatische Kalibrierung deutlich vereinfacht werden, da kein Parameter, zumindest bei kleinen Änderungen, mehrere Szenarien gleichzeitig beeinflussen kann. Insgesamt wurden auf diese Weise 96 entsprechende k_{cat} -Werte gefunden.

Mit den 96 ausgewählten k_{cat} -Werten wurde daraufhin die schon beschriebene automatische Kalibrierung mit einer maximalen erlaubten relativen k_{cat} -Änderung von 50 (entsprechend einer maximal 50-fachen Erniedrigung bzw. Erhöhung) durchgeführt. Die Kalibrierung wurde nach einem Standardlimit von 3.000 Optimierungsrounds durch `fmincon` beendet.

Die aus dieser Kalibrierung erhaltenen k_{cat} -Werte, deren Änderung alle 96 ausgewählten Werte betraf, wurden auf *iJO1366**C angewendet. Das daraus resultierende finale kalibrierte Modell heißt *iJO1366**.

3.4.2 Resultate mit dem kalibrierten Modell

Wachstumsraten unter verschiedenen Kohlenstoffquellen

Zunächst wurden die maximal möglichen Wachstumsraten von *iJO1366** mit den 25 *in vivo* gemessenen Wachstumsraten verglichen (Abb. 3.3). 23 dieser Wachstumsraten stammen aus entsprechend vielen Kohlenstoffquellen (ohne Glucose) unter aeroben Bedingungen und wurden aus [12] entnommen. Hinzu kamen die schon genannten zwei Messungen mit Glucose aus [93], wovon eine unter aeroben und eine unter anaeroben Bedingungen gemessen wurde. Beim damit durchgeführten Vergleich zeigte sich eine hohe Korrelation zwischen den vorhergesagten und den gemessenen Wachstumsraten, mit einem Pearson-Korrelationskoeffizienten von 0,93 und einem Spearman-Korrelationskoeffizienten von 0,91. Im Vergleich dazu erzielte das MOMENT-Modell von *iJO1366* aus der initialen MOMENT-Veröffentlichung mit 24 der 25 hier genutzten Wachstumsszenarien nur einen Pearson- sowie Spearman-Korrelationskoeffizienten von 0,47 [12].

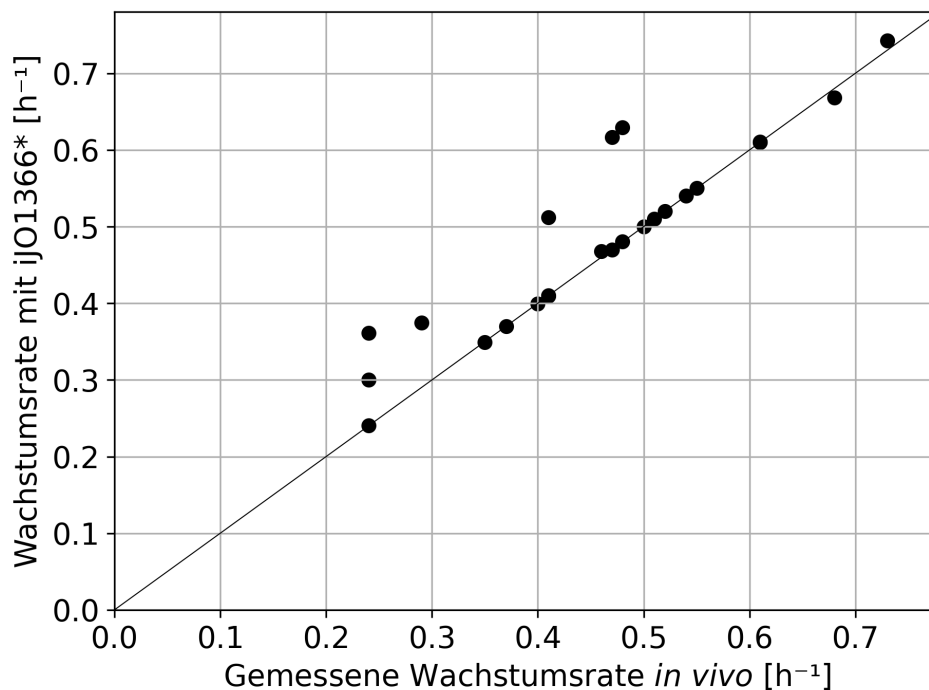


Abbildung 3.3: Streudiagramm der durch *iJO1366** vorhergesagten sowie der in [12] sowie [93] *in vivo* gemessenen Wachstumsraten für die genannten 25 unterschiedlichen Wachstumsszenarien.

Sekretionsverhalten unter Glucose

Zusätzlich zu den im vorherigen Abschnitt dargestellten Vorhersagen der maximalen Wachstumsrate wurde ein Vergleich der vorhergesagten mit *in vivo* gemessenen Austauschraten aus [93] für aerobes und anaerobes Wachstum auf Glucose durchgeführt (Abb. 3.4). Konkret untersucht wurden die Kohlenstoff-Substrat-Aufnahme und die Ausscheidung der wichtigsten Sekretionsprodukte. Hierbei wurde von substratlimitiertem Wachstum ausgegangen, das die Aufnahme von Glucose begrenzt. Das Modell wurde mit verschiedenen Glucose-Aufnahmeraten simuliert, ausgehend von der minimalen (aerobes Wachstum: $0,14 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$, anaerobes Wachstum: $1,26 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$; diese Flüsse sind für die Produktion einer Mindestmenge an ATP für die ATPM-Erhaltungsreaktion notwendig) bis zur maximal möglichen (aerob: $13,83 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$, anaerob: $24,99 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$) Substrat-Aufnahmerate. Bei jeder simulierten Austauschrate wurden dabei die daraus resultierenden Austauschraten für jede Aufnahme bei maximaler Wachstumsrate bestimmt.

Für aerobe Bedingungen wurde festgestellt, dass *iJO1366** bis zum Erreichen einer kritischen Glucose-Aufnahmerate, bei deren Überschreitung Acetat ausgeschieden wird, einen vollständig respiratorischen Stoffwechsel (ohne Ausscheidung kohlenstoffhaltiger Produkte außer CO_2) aufweist. Im Gegensatz zu *iJO1366* und ohne Hinzunahme weiterer Nebenbedingungen, wie etwa eine begrenzte Sauerstoffaufnahme, konnte der bekannte Überlaufstoffwechsel von *E. coli* somit allein durch die Einbeziehung enzymatischer Nebenbedingungen in *iJO1366** erklärt werden, ähnlich zur entsprechenden Ansicht in [66]. Eine sehr gute Übereinstimmung der vorhergesagten Wachstums- und Acetatausscheidungsdaten mit den gemessenen Flussraten aus [93] zeigte sich auch für eine entsprechende Glucose-Aufnahmerate von $9,53 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$.

Anschließend wurden die gleichen Simulationen für anaerobes Wachstum mit unterschiedlichen Glucose-Aufnahmeraten durchgeführt. Die Analyse mit *iJO1366** sagte hierbei für einen weiten Bereich von Substrat-Aufnahmeraten eine, im Vergleich zu anderen Ausscheidungen, dominierende Ausscheidung von Ethanol, Formiat und Acetat als Sekretionsprodukte voraus, was mit biologischen Erkenntnissen übereinstimmt [67]. Die maximale anaerobe Ausbeute von 2,5 ATP pro Glucosemolekül wird durch den gleichzeitigen Ablauf dieser sekretierenden Stoffwechselwege erreicht. Bei einer Substrat-Aufnahmerate von $16,69 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ stimmen die vorhergesagten Austauschraten gut mit den *in-vivo*-Messungen aus [93] überein. Interessanterweise werden mit *iJO1366**, für unter sehr hohen Glucose-Aufnahmeraten, erhöhte Lactat-Produktionsraten, und dazu reduzierte Raten für alle anderen organischen Endprodukt. Die Erhöhung der maximalen Wachstumsrate durch dieses veränderte Sekretionsverhalten wird jedoch nur als marginal angesehen und ist daher *in vivo* wahrscheinlich nicht relevant. Dennoch wurde in [94] *in vivo* gezeigt, dass unter Bedingungen mit hohen Flüssen im zentralen Stoffwechsel die Lactatsynthese zum bevorzugten Sekretionsweg werden könnte, möglicherweise aufgrund der geringeren Proteinkosten im Vergleich zum kombinierten Ablauf der Ethanol-, Acetat- und Formiatsekretion.

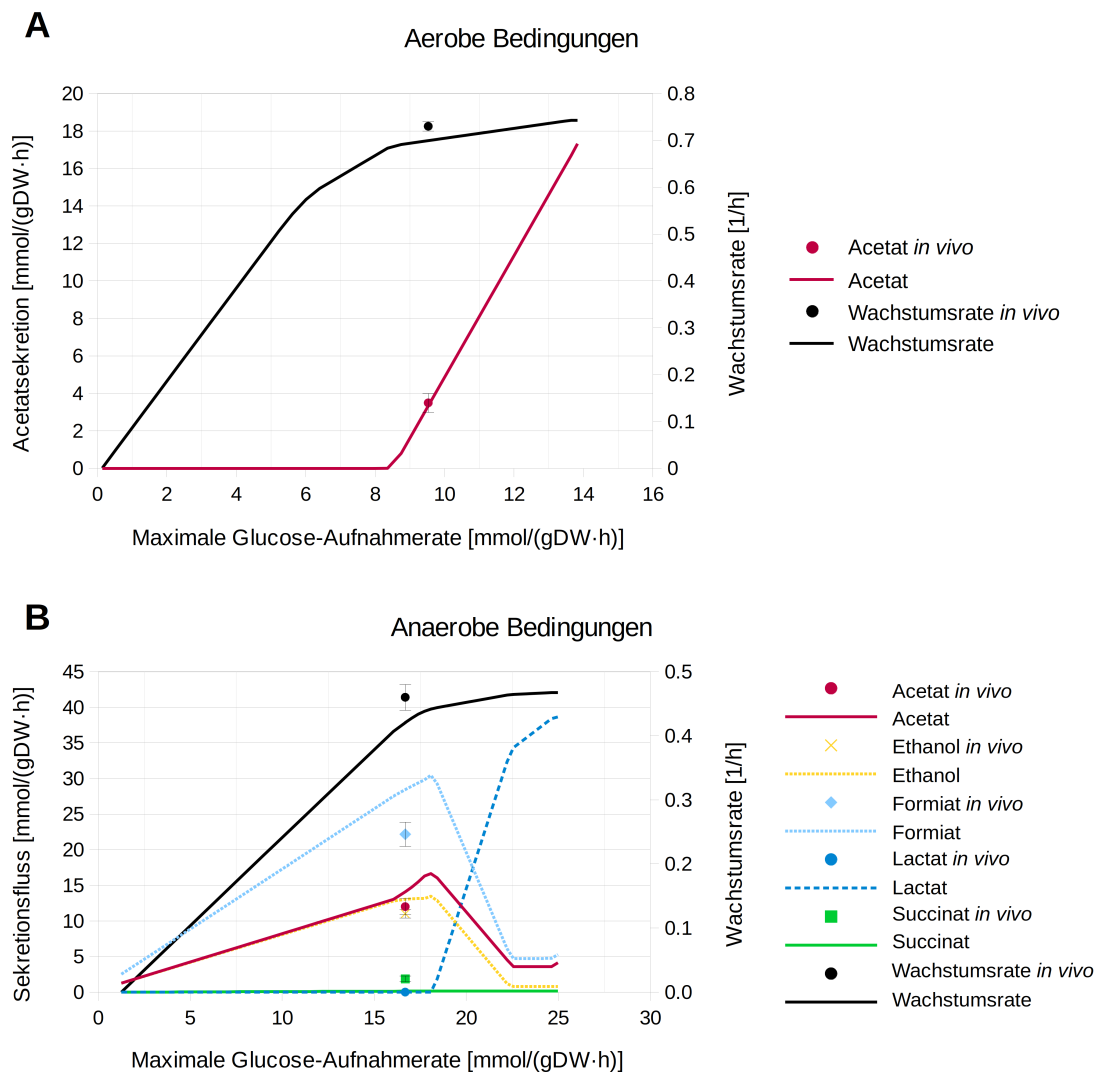


Abbildung 3.4: Vorhergesagte Austauschflüsse von *iJO1366** für den gesamten Bereich möglicher Glucose-Aufnahmeraten und A) aeroben und B) anaeroben Bedingungen. Eine Flussvariabilitätsanalyse zeigte lediglich vernachlässigbare Variabilitäten beim gezeigten Sekretionsverhalten. Zusätzlich werden *in vivo* gemessene entsprechende Flüsse aus [93] mit den entsprechenden Standardabweichungen dargestellt. In dieser Abbildung ist der gelbe Datenpunkt der Ethanol-Sekretion in B direkt unter dem roten Datenpunkt der Acetat-Sekretion, entsprechend liegt auch ein Großteil der gelben Linie unter der roten.

Variabilitätsanalyse

Als weiteres Maß zum Vergleich von *iJO1366* und *iJO1366** wurde eine FVA-basierte Variabilitätsanalyse durchgeführt. Hierfür wurde zunächst eine *iJO1366*-Variante erzeugt, in der alle reversiblen Reaktionen in eine Vor- und Rückwärtsreaktion aufgeteilt sind, um somit die gleiche Anzahl entsprechender Reaktionen in dieser *iJO1366*-Variante wie in *iJO1366** zu erzielen. Danach wurden das schon genannte Standard-Austauschscenario für beide Modelle und eine maximale Glucose-Aufnahmerate von $9,53 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ gesetzt. Mit diesen Einstellungen wurde daraufhin eine FVA aller Reaktionen in der *iJO1366*-Variante sowie *iJO1366** durchgeführt.

Aus der resultierenden absoluten Differenz der berechneten minimalen sowie maximalen Werte wurde die Flussvariabilität einer jeden Reaktion berechnet.

Die kumulative Verteilung der so berechneten Variabilitäten (Abb. 3.5) zeigte, dass *iJO1366** insgesamt eine deutlich geringere Variabilität aufweist als *iJO1366*. Tatsächlich hatte jede Reaktion in *iJO1366** eine entweder gleich große oder kleinere Variabilität als in *iJO1366*. Beispielsweise erreichten in *iJO1366* 81 Reaktionen den künstlich festgelegten maximalen Reaktionsfluss von $1000 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$, während es in *iJO1366** nur 3 Reaktionen waren. Diese Ergebnisse zeigten somit, dass die Einführung enzymatischer Nebenbedingungen zu einer drastischen Verkleinerung des Lösungsraumes möglicher Flussraten führen kann.

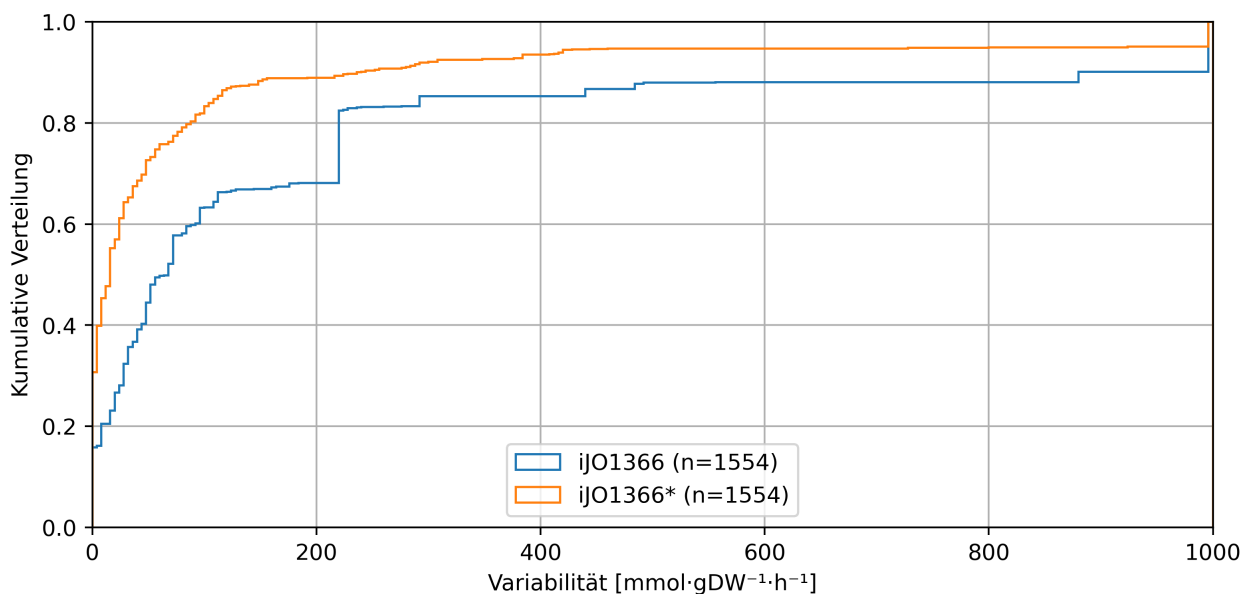


Abbildung 3.5: Vergleichende kumulative Verteilung der Flussvariabilitäten in *iJO1366* sowie *iJO1366**, beide jeweils mit irreversibel aufgeteilten reversiblen Reaktionen, unter aerobem Wachstum mit einer maximalen Glucose-Aufnahmerate von $9,53 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$. Reaktionen mit einem maximalen Fluss von $0 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ sind aus dieser Abbildung ausgeschlossen.

Minimal Cut Sets mit *iJO1366* sowie *iJO1366**

Als weitere Analyse wurde das Konzept der MCS (siehe Gleichungen (2.32) bis (2.33)) verwendet, um gentechnische Interventionen mit und ohne enzymatischen Nebenbedingungen zu berechnen und zu vergleichen. Als Anwendungsbeispiel wurden MCS mit bis zu 6 Interventionen (in Form von Knock-Outs von Reaktionen) für die wachstumsgekoppelte Produktion der Grundchemikalien Ethanol und Succinat sowie für die Aminosäuren Leucin und Valin betrachtet. Dies erfolgte jeweils in *iJO1366* (mit, wie in der vorherigen Analyse, irreversibel aufgespaltenen reversiblen Reaktionen) und in *iJO1366**. Die MCS wurden für anaerobe Bedingungen mit den folgenden Einschränkungen berechnet: Die maximale Glucose-Aufnahmerate in *iJO1366* wurde auf $15 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ festgelegt, während diese Rate in *iJO1366** nicht begrenzt wurde. In beiden Modellen wurden die Austauschreaktionen für alle Kohlenstoffmetabolite deaktiviert, mit Ausnahme der Standardsekretionsprodukte (Acetat, Ethanol, Formiat, Succinat, Lactat, CO_2) und des jeweiligen

Zielprodukts. Für jedes Zielprodukt wurde unabhängig von der Wachstumsrate, als Desired Region (welches auch nach Einbau eines MCS weiter möglich sein muss), eine minimale Wachstumsrate von $0,1 \text{ h}^{-1}$ und, als Target Region (die nach Durchführung der berechneten Interventionen nicht mehr möglich sein soll), eine spezifische maximale Produktausbeute (1,4 mol/mol für Ethanol, 1,0 mol/mol für Succinat, 0,2 mol/mol für Leucin und 0,3 mol/mol für Valin) gefordert. Diese MCS-Einstellungen folgen dem Konzept der starken Koppelung von Wachstum mit der Synthese eines Produktes [29].

Eine Übersicht der so berechneten MCS und deren Verhältnisse untereinander findet sich in Tab. 3.4. Grundsätzlich ist ein sehr heterogenes Bild der MCS-Anzahl und der Verhältnisse der MCS untereinander, je nach Modell sowie Zielprodukt, erkennbar.

So ergab sich mit Ethanol als Zielprodukt, dass die enzymatischen Nebenbedingungen in *iJO1366** eine signifikant höhere Anzahl möglicher MCS-Strategien verursachten (58 % mehr MCS in *iJO1366** im Vergleich zu *iJO1366*). Ein Vergleich der MCS mit den beiden Modellen zeigte, dass etwa ein Viertel der 7.168 MCS in *iJO1366** mit *iJO1366* identisch war. Der größte Anteil (ca. 60 %) der MCS stellte jedoch Strategien dar, die in *iJO1366* nicht existieren, dabei auch nicht als Ober- oder Teilmenge der dort berechneten MCS. Zudem war das Resultat interessant, dass 231 MCS in *iJO1366** errechnet wurden, die Teilmengen von (1.516) MCS in *iJO1366* waren. Die geringere Zahl der erforderlichen Eingriffe in diesen MCS deutete darauf hin, dass die Ethanolsekretion bereits bis zu einem gewissen Grad durch die enzymatischen Nebenbedingungen erzwungen wurde. Andererseits gab es einige solcher Fälle auch in der anderen Richtung, wo (11) MCS von *iJO1366* Teilmengen von (101) MCS in *iJO1366** waren. Ähnliche Ergebnisse ergaben sich für Succinat als Zielprodukt, wobei dort der Anteil der identischen MCS in beiden Modellen größer war.

Ein im Vergleich zu Ethanol sowie Succinat sich unterscheidendes Bild zeigte sich bei den Aminosäuren Leucin sowie Valin. So wurde für Leucin mit *iJO1366** kein einziger MCS gefunden, während mit *iJO1366* mindestens 196 MCS berechnet werden konnten. Eine mögliche Erklärung ist, dass die durch MCS in *iJO1366* erzwungenen Wege für die Leucinsynthese in *iJO1366** aufgrund von Einschränkungen durch die enzymatischen Nebenbedingungen nicht möglich sind. Mithilfe einer Ausbeuteoptimierung konnte festgestellt werden, dass es grundsätzlich möglich ist, die in den MCS-Einstellungen gegebene Leucinausbeute im *iJO1366** unter der gegebenen minimalen Wachstumsrate zu erreichen. Dennoch konnte, wie gezeigt, keine Koppelung mit maximal 6 Knock-Outs erzwungen werden. Im Falle von Valin konnten zwar 3712 MCS in *iJO1366** identifiziert werden, was aber im Vergleich zu *iJO1366* (29290) eine deutlich reduzierte Anzahl darstellte. Wichtig ist dabei, dass 3664 MCS zwar in beiden Modellen identisch waren, aber kein einziger MCS, der nur in einem der beiden Modelle existierte, eine Teilmenge des anderen war. Dies deutete darauf hin, dass in beiden Modellen stark unterschiedliche Strategien auftreten können.

Die Ergebnisse dieser MCS-Berechnungen zeigten somit, dass die Anwendung von sMOMENT zu neuen biotechnologischen Strategien zum Stoffwechseldesign führen kann, die ohne enzymatische Nebenbedingungen nicht gefunden worden wären.

Tabelle 3.4: Vergleichende Resultate mit den MCS die für verschiedene Zielprodukt-Synthesen für iJO1366 sowie iJO1366* errechnet wurden. Die angegebenen Prozentsätze in den Teil- und Obermengen-Angaben beziehen sich auf die Gesamtanzahl berechneter MCS.

Produkt:	Ethanol		Succinat		Leucin		Valin	
Modell:	iJO1366	iJO1366*	iJO1366	iJO1366*	iJO1366	iJO1366*	iJO1366	iJO1366*
# alle berechneten MCS	4.538	7.168	7.801	9.619	196	0	29.290	3.712
# MCS bis Größe 3	0	0	0	0	0	0	0	0
# MCS bis Größe 4	87	189	135	215	24	0	240	0
# MCS bis Größe 5	678	871	1.918	2.148	32	0	3.100	48
# MCS bis Größe 6	3.773	6.108	5.748	7.196	140	0	25.950	3.664
# MCS die Teilmenge der MCS des anderen Modell sind	11 (0,2 %)	231 (3,2 %)	21 (0,2 %)	174 (1,8 %)	0	0	0	0
# MCS die Obermenge der MCS des anderen Modell sind	1.516 (33,4 %)	101 (1,4 %)	1.218 (15,6 %)	42 (0,4 %)	0	0	0	0
# MCS die in beiden Modellen vorkommen	1.899 (41,8 %)	1.899 (26,5 %)	6.141 (78,7 %)	6.141 (63,9 %)	0	0	3.280 (11,2 %)	3.280 (88,4 %)
# MCS die weder identisch noch Teil- oder Obermenge der MCS des anderen Modells sind	1.112 (24,6 %)	4.937 (68,9 %)	421 (5,4 %)	3.262 (33,9 %)	196 (100 %)	0	26.010 (88,8 %)	432 (11,6 %)

3.5 Diskussion

In diesem Kapitel wurden drei wesentliche Entwicklungen vorgestellt: Zunächst wurde die sMOMENT-Methode zum rationalisierten Einbau enzymatischer Nebenbedingungen in metabolischen Modellen entwickelt. Hierzu wurde die AutoPACMEN-Toolbox entwickelt, welche die automatische Konstruktion und Kalibrierung von sMOMENT-Modellen ermöglicht. Schließlich wurde AutoPACMEN verwendet, um die sMOMENT-Version *iJO1366** des genomskaligen *E. coli*-Modells *iJO1366* zu konstruieren. Ein Vergleich dieser beiden Modelle zeigte, wie sich die Nebenbedingungen von sMOMENT auf die wichtigsten Modelleigenschaften und Vorhersagen auswirkten.

sMOMENT verfolgt den gleichen grundlegenden Ansatz wie die schon beschriebene Methode MOMENT [12], siehe Abschnitt 2.1.5, jedoch auf eine rationalisierte Weise. Dabei bestehen drei wesentliche Unterschiede zu MOMENT: (1) Ohne einzubauende gemessene Konzentrationen benötigt sMOMENT keine expliziten Variablen für die Enzymkonzentration. (2) sMOMENT rationalisiert den Umgang mit Isozymen, welche die gleiche Reaktion katalysieren, indem es sich dabei für die schwächste resultierende Nebenbedingung entscheidet, hierbei das Enzym mit dem geringsten Bedarf an dem Verhältnis aus dem Molekulargewicht des Enzyms zum k_{cat} . Dies wiederum hat jedoch keine Auswirkung auf typische Optimierungen mit dem Modell. (3) Enzymatische Nebenbedingungen werden direkt (durch Hinzunahme von nur einem Pseudometaboliten und einer Pseudoreaktion) in die stöchiometrische Matrix constraint-basierter Stoffwechselmodelle integriert. Dies erleichtert die Analyse und Simulation der Modelle mit Programmpaketen der constraint-basierten Modellierung sowie den Export und Import als SBML-Modelle.

Eine, schon in Abschnitt 2.1.5 beschriebene, weitere mit MOMENT und sMOMENT verwandte Methode ist GECKO [13], bei der die relevanten Enzyme sowie ihre Erzeugung und Verwendung ebenfalls als einzelne Metabolite bzw. Reaktionen in die stöchiometrische Matrix eines constraint-basierten Modells einbezogen werden. Eine zentrale Motivation für diese explizite Darstellung in GECKO war die Möglichkeit, gemessene Enzymkonzentrationen direkt in die Matrix zu integrieren und damit das Austauschen von GECKO-Modellen erheblich zu erleichtern. Die Flexibilität von GECKO hat jedoch den schon erwähnten Nachteil, dass GECKO-Modelle besonders umfangreich ausfallen können. So enthält das vollständig erweiterte GECKO-Modell für *iJO1366* (erstellt mit AutoPACMEN, wobei allen Enzymen Pseudo-Konzentrationsmessungen unterzogen wurden) 7728 Reaktionen und 4166 Metabolite, was eine enorme Vergrößerung im Vergleich zu den 3178 Reaktionen und 1806 Metaboliten im sMOMENT-Modell *iJO1366** (vgl. Tab. 3.1) darstellt. Komplexe Analysen, wie etwa die Aufzählung von Minimal Cut Sets, können in einem Netzwerk dieser Größe schwierig oder sogar unmöglich werden. Zudem wurden auch die Flussvorhersagen vor *iJO1366*B* (also *iJO1366** bevor die k_{cat} -Werte mit den Modellkalibrierungen angepasst wurden) mit der entsprechenden GECKO-Version des *iJO1366*-Modells verglichen, wobei jeweils der gleiche maximale Gesamt-Enzympool von $0,095 \text{ g} \cdot \text{gDW}^{-1}$ verwendet wurde. Hierbei fanden sich keine Unterschiede in den Vorhersagen, wenn keine Proteinmessungen vorliegen. Darüber hinaus, und wie schon im Abschnitt 3.2 beschrieben, können auch Enzym-

Konzentrationsmessungen in die Erzeugung eines sMOMENT-Modells einbezogen werden, wobei auch hier, anders als mit GECKO, das Modell so klein wie möglich gehalten wird.

Genauso wie GECKO und MOMENT konzentriert sich sMOMENT auf Nebenbedingungen, welche die gesamte in modellierten Metabolismus relevante Proteinmenge betrachten. Damit ist es konzeptuell deutlich simpler als die verwandten Methoden „Resource Balance Analysis“ [95], kurz RBA, sowie „Metabolism and Expression“ [96], kurz ME. In diesen wird versucht, jegliche Schritte der Genexpression, einschließlich Transkription sowie Translation, explizit zu modellieren. Auch wenn diese Methoden theoretisch eine deutlich bessere Vorhersagekraft erweisen könnten, erfordern sie für realistische Vorhersagen gleichzeitig eine deutlich größere Anzahl *in vivo* oft schwer zu ermittelnder Werte. Gerade für Organismen mit einer großen Menge unbestimmter Parameter bietet sich somit sMOMENT, zusammen mit dessen Implementierung AutoPACMEN, als rationalisierte Methode zum Einbau enzymatischer Nebenbedingungen an.

AutoPACMEN ist zudem, soweit dem Autor bekannt, das erste Programmpaket, welches gleichzeitig das Auffinden aller relevanten enzymatischer Parameter und dazu das Erstellen und das Kalibrieren constraint-basierter Modelle mit enzymatischen Nebenbedingungen ermöglicht. Die GECKO implementierende GECKO Toolbox [83] konnte zum Zeitpunkt der Veröffentlichung von AutoPACMEN nur die ersten beiden Aufgaben durchführen, und ersteres auch nur mit der Datenbank BRENDA [88]. Für MOMENT wurde für alle Aufgaben kein solches Programmpaket veröffentlicht. Für RBA- bzw. ME-Modelle existieren zudem die Programmpaket RBAPy [97] bzw. COBRAME [98]. Diese ermöglichen jeweils zwar eine Erstellung sowie (im Falle von RBAPy) Kalibrierung des Modells mit der entsprechenden Methode, können aber keine weiteren notwendigen Parameter automatisch beziehen. Durch die Implementation von AutoPACMEN sowohl als Kommandozeilenprogramm als auch Python-Modul kann dieses auf einfache Weisen zusammen mit anderen Programmpaketen der constraint-basierten metabolischen Modellierung benutzt werden. Da AutoPACMEN auf COBRAPy [10] basiert, kann es zudem sogar als eine Erweiterung davon angesehen werden. Auch wenn AutoPACMEN in diesem Kapitel nur zur Generierung von *iJO1366** verwendet wurde lässt sich damit auch für jedes andere constraint-basierte Modell, welches UniProt-kompatible Enzym-Reaktions-Regeln und BiGG-Identifizierer für Reaktionen sowie Metabolite beinhaltet, verwenden.

Das um enzymatische Nebenbedingungen angereicherte und daraufhin kalibrierte Modell *iJO1366** hat eine Vielzahl an Anwendungsmöglichkeiten. So ist es, soweit dem Autor bekannt, das erste um enzymatische Nebenbedingungen ergänzte constraint-basierte Modell von *E. coli* welches sowohl unter aeroben als auch anaeroben Bedingungen ein plausibles Sekretionsverhalten aufzeigt. Ferner haben die in diesem Kapitel gezeigten Ergebnisse mit *iJO1366** wesentliche weitere Unterschiede in den Vorhersagen mit *iJO1366* gezeigt. So zeigen die Unterschiede in den vorhergesagten Minimal Cut Sets zu Produktsynthesen (Tab. 3.4) das Potential der Anwendung solcher Modelle im Gebiet des rationalen Designs eines Metabolismus auf.

Auch in sMOMENT verbleiben einige wesentliche Einschränkungen der constraint-basierten Modellierung mit enzymatischen Nebenbedingungen, sowohl MOMENT, GECKO als auch sMOMENT, bestehen. Zum einen lassen sich wesentliche Parameter, wie der Enzympool oder die k_{cat} nur sehr schwer messen und sind in vielen Fällen unbekannt oder sogar je nach

Wachstumsbedingung variabel. Genau betrachtet stellt auch ein *in vivo* oder *in vitro* gemessener k_{cat} letztlich immer nur einen mit einem eigenen, von den Messbedingungen abhängigen, „scheinbaren“ k_{cat} , also einen k_{app} , dar. Die Abhängigkeit eines gemessenen k_{cat} von den Messbedingungen kann dabei stark nicht-linear sein, womit die Nutzung von k_{cat} für unterschiedliche getestete Bedingungen zu stark verzerrten Werten führen kann. Ferner stellt die zentrale Gleichung all dieser Methoden, Gleichung (2.17), nur eine absolut höchste Grenze für den Fluss einer Reaktion dar. Sowohl thermodynamische als auch erweiterte enzymatische kinetische Einflüsse werden nicht berücksichtigt. Die Approximation des Gesamt-Enzymools mithilfe einer mittleren Sättigung der Enzyme (Gleichung (2.19)) stellt zudem eine grobe Vereinfachung dar. Speziell bei der Betrachtung des Überlaufstoffwechsels fehlt sMOMENT, wie auch MOMENT und GECKO, zudem eine Betrachtung des inhibitorischen Einflusses von Sekretionsprodukten wie Acetat [99].

Wie außerdem in [100] festgestellt wurde, gehören enzymatische Nebenbedingungen zusammen mit anderen Ansätzen, wie z.B. einigen thermodynamischen Modellierungsmethoden, zu einer ganzen Klasse constraint-basierter Modellierungsansätze, die oft erfolgreich den Überlaufmetabolismus von Organismen durch ihre jeweiligen neu eingeführten Nebenbedingungen vorhersagen können. Der generelle Mechanismus dahinter ist demnach eine Nebenbedingung, die den Gesamtfluss einschränkt, wie z.B. in sMOMENT der gesamte Enzymool. Welcher Anteil der erfolgreichen Vorhersage des Überlaufstoffwechsels also tatsächlich auf die hergeleiteten enzymatischen Ideen zurückzuführen ist und welcher rein auf den Einfluss des flusslimitierenden Gesamtzymools selbst, muss in weiteren Studien gezeigt werden.

4 Erklärung von NAD(P)(H)-Spezifitäten metabolischer Reaktionen mittels thermodynamischer Modellierung

4.1 Motivation

Die Redoxcofaktoren NAD (Nicotinamid-Adenin-Dinukleotid) und NADP (Nicotinamid-Adenin-Dinukleotid-Phosphat) spielen eine essentielle Rolle als Elektronenträger im Metabolismus von Zellen aller biologischen Reiche. Diese Cofaktoren, die sich nur durch eine Phosphatgruppe unterscheiden, existieren in zwei Formen, und zwar oxidiert (als NAD^+ und NADP^+) sowie reduziert (als NADH und NADPH). Die oxidierten Formen können bei oxidierenden Reaktionen zwei Elektronen aufnehmen, während die reduzierten Formen als Elektronendonatoren in reduzierenden Reaktionen agieren können. Eine kürzlich erschienene Arbeit deutet darauf hin, dass beide Cofaktoren bereits beim letzten gemeinsamen Vorfahren der Bakterien vorhanden waren [101]. Darüber hinaus ist die Spezifität der metabolischen Enzyme, die mit NAD(H) oder NADP(H) zusammenhängen, in den verschiedenen Organismen in Bezug auf die Wahl von NAD(H) oder NADP(H) weitgehend konserviert.

Dieses universelle Vorhandensein sowohl von NAD(H) als auch NADP(H) wirft die Frage auf, warum zwei verschiedene Pools an Redoxcofaktoren, die sich chemisch kaum unterscheiden, verwendet werden. Eine häufig genutzte Erklärung hierfür ist, dass das Vorhandensein von zwei Redoxcofaktor-Pools den gleichzeitigen Ablauf metabolischer Wege mit unterschiedlichen Anforderungen an die thermodynamischen Potentiale der verwendeten Redoxcofaktoren ermöglicht. Obwohl die Standardänderungen der freien Gibbs-Energie zwischen den oxidierten und reduzierten Formen von NAD(H) und NADP(H) (und damit ihre Redoxpotentiale) nahezu zueinander identisch sind (ausgerechnet mittels [102]), unterscheiden sich die tatsächlichen Gibbs-Energien *in vivo* erheblich. Dieser Unterschied ergibt sich daraus, dass das Konzentrationsverhältnis zwischen reduzierten und oxidierten Formen *in vivo* in vielen Spezies, wie etwa *E. coli*, eher niedrig für $\text{NADH}:\text{NAD}^+$ (z.B. ca. 0,02 in *E. coli* [103]), aber deutlich höher für $\text{NADPH}:\text{NADP}^+$ (ca. 30 in *E. coli* [103]) ist. Dieser Unterschied ermöglicht die gleichzeitige Durchführung von Oxidationsreaktionen, begünstigt durch das niedrige $\text{NADH}:\text{NAD}^+$ -Verhältnis, und Reduktionsreaktionen, begünstigt durch das hohe $\text{NADPH}:\text{NADP}^+$ -Verhältnis. Dies wäre mit einem einzigen Cofaktor-Pool vermutlich nicht möglich. Zu dieser Betrachtung passt, dass NAD^+ überwiegend als Elektronenakzeptor bei katabolischen Reaktionen dient, während NADPH in der Regel als Elektronendonator in anabolischen Stoffwechselwegen fungiert. Wenn man jedoch NAD(H) ausschließlich mit dem Katabolismus und NADP(H) ausschließlich mit dem Anabolismus in Verbindung bringt, übersieht man die Notwendigkeit, das verbrauchte NAD^+ und NADPH durch geeignete Reaktionen wiederherzustellen. So wird beispielsweise in heterotrophen Organismen, auf die sich die Arbeit dieses Kapitels konzentriert, NAD^+ in erster Linie durch die Respiration und fermentative Wege regeneriert. Umgekehrt wird der NADPH-Pool häufig über den oxidativen Pentosephosphatweg aufgefüllt, der selbst ein katabolischer Weg ist.

Zwar besteht somit ein allgemeiner Konsens über die funktionelle Rolle von NAD(H) sowie NADP(H), jedoch sind objektive quantitative Konzepte erforderlich, um zu klären, warum die

Nutzung gleich zweier Redoxcofaktor-Pools den Zellen einen evolutionären Vorteil bieten kann. Eine wichtige Arbeit in diese Richtung wurde jüngst von Goldford *et al.* [104] durchgeführt. Die Autoren nutzen dabei constraint-basierte Modellierung ein, um Aspekte wie die thermodynamische Durchführbarkeit von Wachstum in einem metabolischen Netzwerk bei einzelnen Cofaktoraustauschen von NAD(H) zu NADP(H), oder umgekehrt, zu bewerten. Darüber hinaus führten sie kinetische Simulationen an einem kleinen Beispielnetzwerk durch, um zu zeigen, dass die Nutzung mehrerer redundanter Cofaktor-Pools, wie NAD(H) und NADP(H), die Gesamtmenge an Enzymen dank erhöhter thermodynamischer Triebkräfte reduziert werden könnte. Zahlreiche Fragen bleiben jedoch weiterhin unbeantwortet, darunter insbesondere, welche Faktoren die Cofaktor-Spezifität einzelner Reaktionen (und den damit verbundenen Enzymen) innerhalb eines bestimmten Stoffwechselnetzes beeinflussen. Aufgrund des niedrigen NADH:NAD⁺-Verhältnisses in den Zellen würde eine Reaktion, die auf die Reduktion eines Redoxcofaktors abzielt, stärker davon profitieren, wenn sie NAD⁺ verwendet. Doch wie bereits erwähnt, müssen zumindest einige Reaktionen NADP⁺ verwenden, um den NADPH-Pool wieder aufzufüllen. Dies wirft die Frage nach der idealen Verteilung der NAD(P)(H)-Spezifitäten auf der Ebene des Netzwerks auf: Wenn allen Reaktionen die Cofaktor-Spezifitäten frei zugeordnet werden könnten, welche Verteilung wäre dann optimal, bspw. im Hinblick auf die netzwerkweite thermodynamische Triebkraft innerhalb des Netzwerks? Lassen sich dann außerdem die erforderlichen NAD(P)H:NAD(P)⁺-Verhältnisse vorhersagen?

Um diese und weitere Fragen zu beantworten und somit neue Erkenntnisse über redundante Redoxcofaktoren zu gewinnen, wird im Rahmen dieses Kapitels TCOSA (für Englisch *Thermodynamics-based Cofactor Swapping Analysis*, zu Deutsch *Thermodynamik-basierte Cofactor-Austausch-Analyse*) entwickelt. TCOSA ermöglicht die Analyse des Effekts des Austausches von Cofaktoren auf die netzwerkweite thermodynamische Triebkraft eines gegebenen constraint-basierten Modells. Ähnlich zu den Arbeiten in [104] basiert TCOSA auf der Verbindung der constraint-basierten Modellierung mit thermodynamischen Nebenbedingungen, verwendet jedoch als zentralen Unterschied die Definition des netzwerkweiten MDF (Abb. 2.1D) zur Ermittlung der maximalen thermodynamischen Triebkraft eines metabolischen Modells unter gegebenen Bedingungen. Dieser Ansatz benötigt keine kinetischen Parameter und erlaubt die Berechnung verschiedener Effekte bezüglich der Nutzung und des Austausches der Redoxcofaktoren. Im Speziellen ermöglicht der Ansatz etwa die Vorhersage von Cofaktor-Spezifitäten, die den MDF maximieren. Diese Spezifitäts-Verteilung kann wiederum mit der *in vivo* auftretenden Verteilung verglichen werden. Auch weitere Eigenschaften, wie die internen Konzentrationsverhältnisse der Cofaktoraare bei maximalen thermodynamischen Triebkräften, lassen sich so untersuchen.

4.2 Das TCOSA-Framework

TCOSA ermöglicht, für ein gegebenes metabolisches Netzwerk, die systematische Analyse der Effekte variiert NAD(P)(H)-Spezifitäten der Redoxreaktionen, vor allem im Hinblick auf die möglichen netzwerkweiten thermodynamischen Triebkräfte. Die zentrale Methodik von TCOSA umfasst eine spezielle Rekonfiguration des Modells sowie constraint-basierte thermodynamische Berechnungen mit diesem rekonfigurierten Modell.

Rekonfiguration eines Modells für TCOSA

Die Netzwerk-Rekonfiguration für TCOSA wirkt sich auf die NAD(P)(H)-abhängigen Redoxreaktionen innerhalb des Modells aus (Abb. 4.1A). Grundsätzlich wird jede dieser Redoxreaktionen dupliziert, was dazu führt, dass es für jede Reaktion sowohl eine NAD(H)- als auch NADP(H)-abhängige Variante gibt. Durch entsprechende Änderungen an den Identifizierern der Reaktionen lässt sich später leicht feststellen, welche Variante *in vivo* im Wildtyp vorhanden ist und welche durch die Duplikation in das Modell eingeführt wurde. Eine Ausnahme wird für Reaktionen gemacht, an denen gleichzeitig NAD(H) *und* NADP(H) beteiligt sind, wie z.B. Transhydrogenasen. Diese Reaktionen werden nicht dupliziert, um zu verhindern, dass das Modell nicht *in vivo* vorkommende Umwandlungen zwischen NAD und NADP durchführen kann. Diese Modellrekonfiguration ähnelt stark der von OptSwap [105]. Die Anwendung von OptSwap war jedoch auf spezifische Oxidoreduktasen beschränkt und enthielt keine thermodynamischen Nebenbedingungen. Um die Verwendung thermodynamischer Nebenbedingungen in TCOSA zu vereinfachen, werden außerdem alle mit thermodynamischen Daten versehenen reversiblen Reaktionen in irreversible umgewandelt, sodass kein Reaktionsfluss negativ sein kann.

Mit einem solchen für TCOSA rekonfigurierten Modell können vier verschiedene Szenarien an NAD(P)(H)-Spezifitäten definiert werden (Abb. 4.1B):

1. „Wildtyp-Spezifität“: In diesem Szenario sind nur die *in vivo* auftretenden ursprünglichen NAD(P)(H)-Spezifitäten zulässig. Alle nicht *in vivo* auftretenden Reaktionsvarianten werden blockiert, indem der jeweilige Reaktionsfluss auf $0 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ gesetzt wird.
2. „Einzelcofaktor-Spezifität“: Hier werden alle NADP(H)-abhängigen Reaktionsvarianten blockiert, sodass alle entsprechenden Redoxreaktionen NAD(H)-abhängig werden. Da in einer typischen Pseudo-Reaktion des Wachstums NADP(H) als essentieller Teil der Biomasse definiert ist, enthält dieses Szenario zusätzlich eine thermodynamisch nicht limitierte Pseudo-Reaktion, die NAD(H) in NADP(H) umwandelt.
3. „Flexible Spezifität“: Sowohl alle NAD(H)- als auch NADP(H)-abhängigen Varianten einer jeden Reaktion sind erlaubt. Somit kann in einer Berechnung eines Modells mit dieser Spezifität für jede Reaktion frei zwischen NAD(H) sowie NADP(H) entschieden werden. Wie weiter unten formal beschrieben, sorgen zusätzliche Nebenbedingungen dafür, dass nicht mehr als eine NAD(H)- und NADP(H)-Variante *einer* Reaktion gleichzeitig ablaufen kann.
4. „Zufällige Spezifität“: Dies ist eine randomisierte Spezifität, bei der ein stochastischer Münzwurf entscheidet, ob entweder die NAD(H)- oder die NADP(H)-abhängige Variante einer Reaktion aktiv sein darf und damit die jeweils andere Variante blockiert. Bei Anwendungen von TCOSA auf ein Modell von *E. coli* mit dieser Spezifität (siehe nächster Abschnitt) wurden stets 1.000 zufällige Spezifitäten berechnet. 500 davon beinhalteten eine völlig freie Spezifität, während bei den 500 anderen zufälligen Spezifitäten die Anzahl der NAD(H)- bzw. NADP(H)-abhängigen Reaktionen identisch zum Wildtypen sein musste. Alle zufälligen Spezifitäten, die bei der höchsten getesteten Wachstumsrate thermodynamisch unmöglich sind ($\text{MDF} < 0,1 \text{ kJ/mol}$; siehe nächster Abschnitt), werden hier für weitere Analysen nicht berücksichtigt.

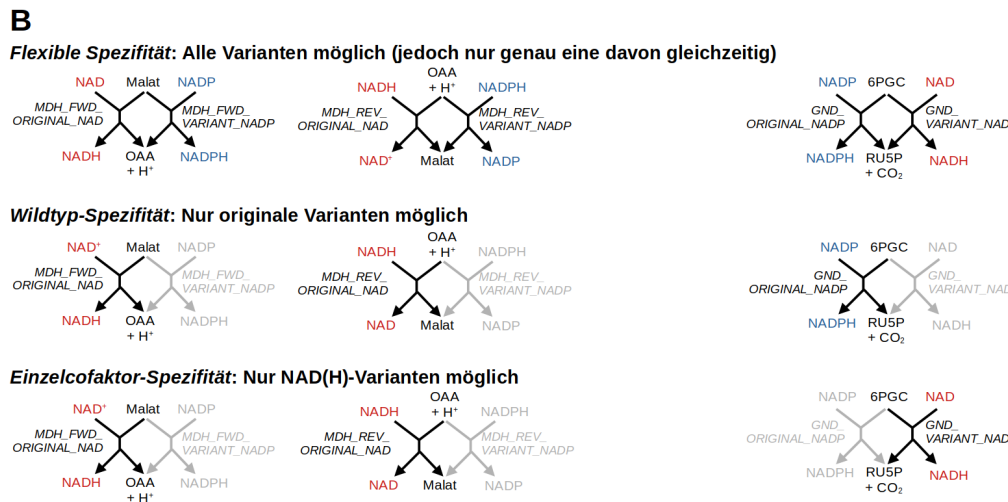
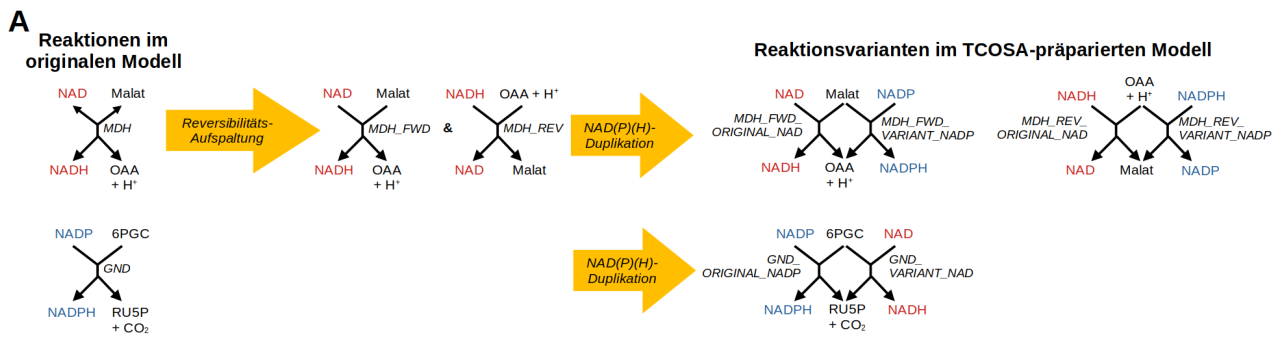


Abbildung 4.1: Modelländerungen in TCOSA und Definition der Spezifitäts-Szenarien. (A) Beispiele der Änderungen um mehrere verschiedene Spezifitäten zu erlauben (MDH: Malat-Dehydrogenase-Reaktion; GND: Phosphogluconat-Dehydrogenase-Reaktion; OAA: Oxalacetat; 6PGC: 6-Phosphogluconat; RU5P: Ribulose-5-Phosphat). (B) Drei der vier in diesem Kapitel betrachteten Cofaktor-Spezifitätsszenarien. Zufälliger Spezifitäten ähneln der Wildtyp-Spezifität, mit dem Unterschied, dass die Wahl der NAD(H)- oder NADP(H)-Spezifität randomisiert ist.

Berechnungsmethoden von TCOSA

Die thermodynamischen Berechnungsmethoden von TCOSA basieren auf OptMDFpathway [15], welches in Abschnitt 2.6 beschrieben wurde.

TCOSA-Ergänzung zu OptMDFpathway

Wie bereits erwähnt, wird in TCOSA für jede Richtungs- und NAD(P)(H)-Variante einer Reaktion eine zusätzliche Nebenbedingung eingeführt, die verhindert, dass mehr als eine Variante dieser Reaktion gleichzeitig aktiv sein kann. Damit soll insbesondere bei der flexiblen Spezifität verhindert werden, dass statt einer direkten NAD(P)(H)-Verteilung nur der Fall des gleichzeitigen Ablaufs mehrerer Cofaktor-Reaktionsvarianten bei der Cofaktorwahl zu Effekten in den Resultaten führt. Um diese zusätzliche Nebenbedingung einzuführen, wird für jede Reaktion k des ursprünglichen Stoffwechselmodells eine Indexmenge A_k gebildet, die alle Indizes der Reaktionsvarianten der Reaktion k enthält. Für jedes A_k mit mehr als einem Element wird nun die folgende Nebenbedingung in Anlehnung an die Formulierungen von OptMDFpathway (Gleichungen (2.25) bis (2.31)) eingeführt:

$$\sum_{i \in A_k} z_i \leq 1 \quad (4.1)$$

Dementsprechend ändert sich die Optimierung des MDF von OptMDFpathway zu:

$$\begin{aligned} & \text{maximiere } B \\ & \quad r, x, z, B \\ & \text{u. d. N. (2.6) bis (2.8), (2.25) bis (2.30), (4.1)} \end{aligned} \quad (4.2)$$

SubMDF

Als neue MDF-basierte Definition wird in TCOSA der SubMDF eingeführt (Abb. 4.2). Dieser bezieht sich im Gegensatz zum (netzwerkweiten) MDF nur auf eine frei wählbare Teilmenge aller Reaktionen eines constraint-basierten Modells. Im Kontext von TCOSA ist dies die Menge aller NAD(P)(H)-abhängigen Reaktionen. Deren SubMDF kann somit unabhängig von anderen thermodynamischen Flaschenhälsen (englisch *bottlenecks*) betrachtet werden kann. Definitionsgemäß ist der SubMDF mindestens so groß wie der MDF.

MDF eines Unternetzwerks (SubMDF): Finde Flussverteilung r welche den MDF ausgewählter Reaktionen maximiert
 z.B., $\text{SubMDF} = \text{MDF}(R_1, R_8)$ innerhalb gegebener **Konzentrationsbereiche** und **Flussnebenbedingungen** und verlangtem **minimalen MDF** (z.B., >0 kJ/mol) aller anderen aktiven Reaktionen

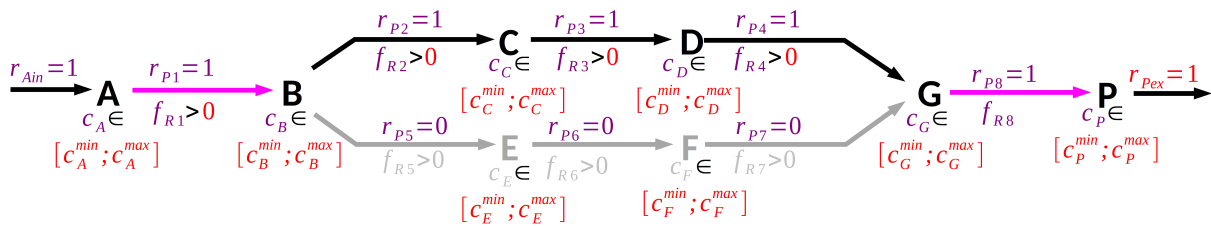


Abbildung 4.2: Auf dem Beispielmotell und den Definitionen von Abb. 2.1 basierende Visualisierung des SubMDF. Rote Farben stehen für gegebene Nebenbedingungen, violette für die im Rahmen der SubMDF-Optimierung zurückgegebenen Werte und Magenta für die für den SubMDF ausgewählten Reaktion.

Die Indizes der im SubMDF betrachteten relevanten Teilmenge von Reaktionen bilden dabei die Menge S . Nun wird für jede Reaktion dieser Menge eine neue Nebenbedingung der Form

$$B_{\text{submdf}} \leq f_j + M \cdot (1 - z_j), \quad \forall j \in S \quad (4.3)$$

eingeführt. B_{submdf} repräsentiert hierbei den SubMDF. Um sinnvolle Werte für den SubMDF zu erhalten, muss eine weitere Nebenbedingung hinzugefügt werden, die sicherstellt, dass die Gesamtlösung thermodynamisch möglich ist (netzwerkweiter $\text{MDF} > 0$ kJ/mol). Dies wird durch eine Nebenbedingung der Form

$$B \geq B_{\text{min}} \quad (4.4)$$

erreicht. B_{min} ist hierbei die minimal zu erreichende netzwerkweite thermodynamische Triebkraft, und wurde im Rahmen dieser Arbeit auf 0,1 kJ/mol gesetzt. Der SubMDF lässt sich nun über folgende Optimierung ermitteln:

$$\begin{aligned} & \underset{r,x,z,B}{\text{maximiere}} B_{\text{submdf}} & (4.5) \\ & \text{u. d. N. (2.6) bis (2.8), (2.25) bis (2.30), (4.1), (4.3) bis (4.4)} \end{aligned}$$

Analyse von Metabolit-Konzentrationsverhältnissen

Zur Analyse der relativen Konzentrationsverhältnisse von $[NADH]:[NAD^+]$ (P_{NADH}) sowie $[NADPH]:[NADP^+]$ (P_{NADPH}) wird in TCOSA eine Variabilitätsanalyse von P_{NADH}/P_{NADPH} durchgeführt. Die logarithmierte Variante dieses Verhältnisses lässt sich in einem MILP als Variable ρ mit

$$\rho = \ln\left(\frac{[NADH]/[NAD^+]}{[NADPH]/[NADP^+]}\right) = x_{NADH} + x_{NADP^+} - x_{NADPH} - x_{NAD^+} \quad (4.6)$$

eingeführen. Im Kontext von TCOSA wird dieses Konzentrationsverhältnis bei einem vorher gemessenen (Sub)MDF berechnet, was für einen entsprechenden MDF als B_{max} über

$$B = B_{max} \quad (4.7)$$

und für einen SubMDF als B_{submax} über

$$B = B_{submax} \quad (4.8)$$

erzielt wird. Das Maximum und Minimum von ρ lässt sich nun über die beiden Optimierungen

$$\underset{r,x,z}{\text{maximiere}} \rho \quad (4.9)$$

$$\text{u. d. N. (2.6) bis (2.8), (2.25) bis (2.30), (4.1), (4.3), (4.7) oder (4.8)}$$

sowie

$$\underset{r,x,z}{\text{minimiere}} \rho \quad (4.10)$$

$$\text{u. d. N. (2.6) bis (2.8), (2.25) bis (2.30), (4.1), (4.3), (4.7) oder (4.8)}$$

ermitteln. Das Verhältnis von P_{NADH}/P_{NADPH} ergibt sich hiernach durch die Anwendung der Exponentialfunktion auf die minimalen und maximalen Werte von ρ .

Minimale Cofaktoraustausch-Analyse

Zur Bestimmung der Mindestanzahl an Reaktionen, die ihre im Wildtyp vorhandene NAD(P)(H)-Spezifität mit dem jeweils anderen Cofaktorpaar austauschen müssen, damit der MDF der flexiblen Spezifität erreicht werden kann, wird wie folgt vorgegangen: Zunächst werden alle Indizes von NAD(P)(H)-abhängigen Reaktionen mit zum Wildtyp komplementärer (also in der flexiblen Spezifität, aber nicht im Wildtypen vorkommender) Cofaktor-Spezifität als Menge C gesammelt. Hinzu kommen Binärvariablen der Form $z_{i,tausch} \in \{0,1\}$, $i \in C$. Diese erlauben den Wechsel von der Wildtyp-Spezifität zur komplementären Spezifität durch folgende Nebenbedingung:

$$r_i \leq z_{i, \text{tausch}} \cdot \beta_i, \quad \forall i \in C \quad (4.11)$$

r_i ist hierbei die Flussrate der Reaktion i mit der alternativen Spezifität, welche nur aktiv werden kann, wenn $z_{i, \text{tausch}} = 1$. Um nun die entsprechende minimale Anzahl notwendiger Cofaktoraustausche, ausgehend vom Wildtypen, zu ermitteln, wird das Resultat des folgenden Optimierungsproblems ermittelt:

$$\begin{aligned} & \underset{r, x, z}{\text{minimiere}} \sum z_{i, \text{tausch}} & (4.12) \\ & \text{u. d. N. (2.6) bis (2.8), (2.25) bis (2.30), (4.1), (4.3), (4.7) oder (4.8), (4.11)} \end{aligned}$$

Bestimmung thermodynamischer Flaschenhalse

Wie bereits in den ursprünglichen Arbeiten zum MDF [47] und OptMDFpathway [15] festgestellt wurde, enthält jede MDF-optimale Lösung mindestens einen thermodynamischen Flaschenhals in Form einer Reaktion. Solche Flaschenhalsreaktionen können als Reaktionen definiert werden, bei denen eine Verringerung ihres $\Delta_r G^\circ$ zu einer Erhöhung des MDF führt [15]. Bottleneck-Reaktionen können daher mithilfe eines MILP wie folgt identifiziert werden: Zunächst wird eine neue Menge von Binärvariablen $z_i^b \in \{0, 1\}$ für alle Reaktionen definiert und Gleichung (2.30) zu

$$B \leq f_i + M - z_i \cdot M + M \cdot z_i^b \quad (4.13)$$

modifiziert. Indem nun beispielsweise für den MDF ein minimaler Wert von 0,1 kJ/mol gefordert wird ($B \geq 0,1$ kJ/mol) und die Summe dieser neuen Binärvariablen per

$$\begin{aligned} & \underset{r, x, z, z^b, B}{\text{minimiere}} \sum z_i^b & (4.14) \\ & \text{u. d. N. (2.6) bis (2.8), (2.25) bis (2.30), (4.1), (4.13)} \end{aligned}$$

minimiert wird, kann die minimale Anzahl an Reaktionen ermittelt werden, bei denen eine Erniedrigung des $\Delta_r G^\circ$ zum Erreichen eines MDF von 0,1 kJ/mol führt. Dies liegt rechnerisch darin begründet, dass die Ergänzung $M \cdot z_i^b$ in Gleichung (4.13) die Nebenbedingung für B praktisch aufgehoben wird.

Diese Routine ähnelt der Funktion „relax_dgo“ in pyTFA [106]. Anders als dort wird jedoch hier nicht die minimale Summe der absoluten $\Delta_r G^\circ$ -Änderungen, sondern die minimale Anzahl betroffener Reaktionen gesucht.

Implementation

Alle Berechnungen dieses Kapitels wurden mithilfe von Pythonskripten durchgeführt. Für das Laden und Speichern metabolischer Modelle (in Form des SBML-Formates) und für grundlegende Flussanalysen wurde COBRApy verwendet [10]. Die Bibliothek pulp [107] wurde für die Konstruktion der MILP verwendet, wobei IBM CPLEX 12.10 als Optimierungslöser verwendet wurde. Alle Skripte sowie erstellten Daten und Modelle (im SBML-Format [78]) können, zusammen mit einer Dokumentation, im folgenden GitHub-Repository eingesehen und heruntergeladen werden:

<https://www.github.com/klamt-lab/TCOSA>

Die Berechnungen wurden auf einem Computercluster (mit insgesamt 16 Intel Xeon Silver 4110 CPU-Kernen sowie 192 GB DDR4-RAM pro Einheit) ausgeführt. Das sequentielle Ausführen aller im Folgenden gezeigten Analysen dauerte auf diesem System ca. 6 Tage.

4.3 Anwendung von TCOSA auf *E. coli* und Analyse der Resultate

Vorbereitung des Modells

Als beispielhafte Anwendung wurde TCOSA auf das genomskalige metabolische *E. coli*-Netzwerk *iML1515* angewendet [73]. Dieses Modell wurde entsprechend der schon erläuterten TCOSA-Modellrekonfiguration wie folgt erweitert und modifiziert, um es für die Anwendung mit TCOSA zu rekonfigurieren:

- Für die thermodynamischen Berechnungen wurden Standard-Gibbs-Energien mithilfe der Python-Programmierschnittstelle des eQuilibrator-Paketes [45] bestimmt. Als Parameter wurden ein pMg von 2,5, eine Ionenstärke von 250 mM, ein cytosolischer pH von 7,5 und, für Multikompartimentreaktionen, eine Standard-Membranpotentialdifferenz von 150 mV bei einem ΔpH von 0,5 zwischen Cytoplasma und Periplasma sowie Periplasma und extrazellulärem Kompartiment verwendet. Für die Wachstums-Pseudoreaktion und alle 331 Austausch-Pseudoreaktionen wurde kein $\Delta_r G^\circ$ festgelegt, wodurch sie thermodynamisch uneingeschränkt blieben. In den verbleibenden $2712-332=2380$ Reaktionen in *iML1515* konnte für 1683 Reaktionen ein $\Delta_r G^\circ$ erfolgreich berechnet werden. Für die verbleibenden 697 Reaktionen, für die der eQuilibrator keine Lösung fand, wurde ein für metabolische Reaktionen sehr niedriger $\Delta_r G^\circ$ von -100 kJ/mol gesetzt, um zumindest eine minimale Einschränkung der effektiv möglichen assoziierten Metabolitkonzentrationen zu erreichen. Auf die gleiche Weise wurde bei den 611 multikompartimentalen Reaktionen verfahren, deren Rolle rein als Transporter definiert war oder bei denen keine Metabolitenumsetzung stattfindet, da häufig nur die Anwesenheit, nicht aber der Mechanismus der Transporter sicher bekannt ist. Eine Ausnahme bildeten multikompartimentale Reaktionen wie die ATP-Synthase und die NADH-Dehydrogenase, bei denen der Protonentransport mit einer Stoffumwandlung gekoppelt ist. Darüber hinaus wurden NAD(P)(H)-abhängige Reaktionen aufgrund ihrer Bedeutung für die vorliegende Arbeit gesondert behandelt: Während für 161 der 238 (68 %) solcher Reaktionen ein $\Delta_r G^\circ$ berechnet werden konnte, konnte für 77 Reaktionen kein $\Delta_r G^\circ$ zugeordnet werden. Um für diese einen realistischeren Standardwert für den $\Delta_r G^\circ$ zu erhalten, wurde der Median der $\Delta_r G^\circ$ aller Reaktionen berechnet, in denen NAD^+ (+15,77 kJ/mol), NADH (-15,79 kJ/mol), NADP^+ (+15,33 kJ/mol) oder NADPH (-15,33 kJ/mol) als Substrat vorkommt. Je nach Verwendung des entsprechenden Cofaktors wurde für die 77 Reaktion der entsprechende Median verwendet.
- Für die Metabolitkonzentrationen wurde, wie in der Originalpublikation von OptMDFpathway [15], ein Bereich von 10^{-6} M bis 0,02 M festgelegt. Für Wasser und Protonen wurde eine Konzentration von 1 M angesetzt, da deren Konzentrationen schon in den eQuilibrator-Berechnungen der $\Delta_r G^\circ$ eingeflossen sind.

- Mit den so gesammelten thermodynamischen Parametern wurde die TCOSA-Rekonfiguration wie im letzten Abschnitt beschrieben (Abb. 4.1A) auf *iML1515* angewendet. Dabei wurde bei allen auftretenden Rückwärtsreaktionen von Reaktionen, denen der eQuilibrator einen $\Delta_r G^\circ$ zuordnen konnte, der negierte $\Delta_r G^\circ$ der Vorwärtsrichtung zugeordnet.

Das aus diesen Daten und Änderungen resultierende Modell wurde *iML1515_TCOSA* genannt. Es besteht aus insgesamt 3982 Reaktionen und 1881 Metaboliten. 299 dieser Reaktionen verwenden NAD^+ und NADH , weitere 299 Reaktionen verwenden NADP^+ und NADPH , während unverändert 6 Reaktionen gleichzeitig NAD(H) und NADP(H) verwenden.

Zusätzlich zu *iML1515_TCOSA* wurde eine Variante mit drei Redoxcofaktorpaaren mit der Bezeichnung *iML1515_3TCOSA* erstellt. Hierbei wurden die NAD(P)(H) -abhängigen Reaktionen nicht nur verdoppelt, sondern verdreifacht, was auch in den Reaktions-Identifizierern zum Ausdruck kommt. Das dritte Cofaktorpaar wurde über die hypothetischen Metabolite NADZ sowie NADZH repräsentiert.

Sowohl *iML1515_TCOSA* als auch *iML1515_3TCOSA* erlauben die gleichen Standard-Austauschreaktionen wie das Originalmodell *iML1515* und verwenden weiterhin eine ATPM-Erhaltungs-Pseudoreaktion mit einer Flussrate von mindestens $6,86 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$. Mit Glucose als Kohlenstoffsubstrat wurde in allen Berechnungen eine maximale Aufnahmerate von $10 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ (aerob) bzw. $20 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ (anaerob) angenommen. Für Acetat als weiteres beispielhaftes Kohlenstoffsubstrat, mit dem im gegebenen Standardmedium von *iML1515* nur aerobes Wachstum möglich ist und daher nur aerob betrachtet wurde, wurde eine maximale Aufnahmerate von $10 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ verwendet.

Mittels einer Flussbilanzanalyse (Gleichungen (2.6) bis (2.8)) wurde zunächst die maximale Wachstumsrate (μ_{\max}) von *iML1515_TCOSA* auf Glucose ohne thermodynamische Nebenbedingungen berechnet. Mit der Wildtyp-Spezifität der Redoxcofaktoren (Abb. 4.1B) ergab sich ein μ_{\max} von $0,877 \text{ h}^{-1}$ unter aeroben respektive $0,375 \text{ h}^{-1}$ unter anaeroben Bedingungen, was erwartungsgemäß den Werten mit *iML1515* entspricht. Interessanterweise wurden jedoch höhere μ_{\max} von $0,881 \text{ h}^{-1}$ (aerob) bzw. $0,47 \text{ h}^{-1}$ (anaerob) mit der Einzelcofaktor-Spezifität beobachtet. Dies deutet darauf hin, dass der Einsatz *eines* Cofaktor-Paares stöchiometrisch effizienter ist, auch wenn *in vivo* gravierende Effekte zu erwarten sind, wie auch die nachfolgenden Ergebnisse auf thermodynamischer Ebene zeigten. Um numerische Probleme zu vermeiden, wurde für alle folgenden Berechnungen die maximale Wachstumsrate auf 99 % von μ_{\max} unter der Wildtyp-Spezifität festgelegt. Dies gilt auch für den unter Acetat entsprechend ermittelten μ_{\max} von $0,21 \text{ h}^{-1}$.

Mit der so festgelegten maximalen Wachstumsrate ergibt sich bspw. aerob unter Glucose ein extrem negativer MDF von ca. -4705 kJ/mol , was daher eine Analyse der thermodynamischen Flaschenhalse erforderlich macht. Die Anwendung der beschriebenen Methode zur Identifizierung der Mindestanzahl von Reaktionen, deren $\Delta_r G^\circ$ für ein thermodynamisch mögliches Wachstum mit entsprechendem μ_{\max} gelockert werden muss, ergab sowohl unter Glucose (aerob wie anaerob) als auch unter Acetat, dass 9 bestimmte Reaktionen thermodynamische Flaschenhälse darstellen (Tab. 5.1). All diese Reaktionen treten im Anabolismus auf und enthalten oftmals extreme $\Delta_r G^\circ$ -Werte.

Für alle weiteren Berechnungen werden die in der Tabelle angegebenen entsprechend modifizierten $\Delta_r G^\circ$ verwendet.

Tabelle 4.1: Liste der Reaktionen mit gelockerten $\Delta_r G^\circ$ -Werten, um in den vorliegenden Berechnungen Wachstum mit iML1515_TCOSA thermodynamisch möglich zu machen. Der $\Delta_r G^\circ$ von AIRC3 wurde dabei auf +100 kJ/mol gesetzt da dessen Rückwärtsrichtung einen thermodynamischen Flaschenhals darstellt. Alle sonstigen Reaktionen sind irreversibel.

BiGG-ID der Reaktion	Name des Enzyms der Reaktion	Durch eQuilibrator ermittelter $\Delta_r G^\circ$ [kJ/mol]	Korrigierter $\Delta_r G^\circ$ [kJ/mol]
KDOCT2	3-Deoxy-Manno-Octulsonat-Cytidyltransferase	+225,43	-100
MECDPS	2-C-Methyl-D-Erythritol 2,4-Cyclodiphosphat Synthase	+1809,64	-100
DHPPDA2	Diaminohydroxyphosphoribosylaminopyrimidin-Deaminase	+278,17	-100
ATPPRT	ATP-Phosphoribosyltransferase	+4125,05	-100
IG3PS	Imidazole-glycerol-3-Phosphat-Synthase	+233,36	-100
MCTP1App	Murein quervernetzende Transpeptidase	+269,73	-100
MALCOAMT	Malonyl-CoA-Methyltransferase	+55,9	-100
AIRC3	Phosphoribosylaminoimidazol-Carboxylase	-30,08	100
SHCHD2	Sirohydrochlorin-Dehydrogenase	+2432,76	-100

Vergleich maximaler thermodynamischer Triebkräfte mit verschiedenen Szenarien an NAD(P)(H)-Spezifitäten

Das beschriebene Modell iML1515_TCOSA wurde nun verwendet, um die maximal erreichbaren netzwerkweiten thermodynamischen Triebkräfte (MDF sowie SubMDF) mit den genannten Szenarien an NAD(P)(H)-Spezifitäten (siehe Abschnitt 4.2) zu vergleichen. Hierbei diente, wie in allen folgenden Analysen in denen es nicht explizit anders genannt wird, Glucose als Kohlenstoffsubstrat. Dabei wurden für jede Spezifität sowohl aerobe als auch anaerobe Bedingungen betrachtet, ausgehend von der wie beschrieben ermittelten maximalen Wachstumsrate, die dann in Schritten von $0,05 \text{ h}^{-1}$ bis zum Erreichen von $0,05 \text{ h}^{-1}$ reduziert wurde. Somit wurden unter aeroben Bedingungen 18 Wachstumsraten ($0,868 \text{ h}^{-1} \dots 0,05 \text{ h}^{-1}$) und unter anaeroben Bedingungen 8 Wachstumsraten ($0,371 \text{ h}^{-1} \dots 0,05 \text{ h}^{-1}$) untersucht.

Die Resultate sind in Abb. 4.3 zusammengefasst. Die (theoretisch optimale) flexible Spezifität erreichte sowohl unter aeroben (Abb. 4.3A), als auch unter anaeroben Bedingungen (Abb. 4.3C) einen MDF von ca. 8 kJ/mol. Unter aeroben Bedingungen wurde dieser MDF-Wert bei allen getesteten Wachstumsraten mit Ausnahme der höchsten erreicht. Unter anaeroben Bedingungen hingegen war der MDF bei höheren Wachstumsraten deutlich reduziert und wurde erst bei niedrigeren Wachstumsraten erreicht. Grundsätzlich konnte auch festgestellt werden, dass der MDF mit der flexiblen Spezifität gleichzeitig den höchsten erreichbaren Wert darstellt. Dies ist auf die definitionsgemäße Freiheit der flexiblen Spezifität zurückzuführen, jede NAD(P)(H)-Spezifität nutzen und diese sogar zwischen verschiedenen Wachstumsraten verändern zu können.

Betrachtet man nun die mit der Wildtyp-Spezifität erreichbaren MDF, so zeigte sich, dass der theoretisch optimale MDF der flexiblen Spezifität bei allen niedrigen und mittleren Wachstumsraten exakt erreicht wurde. Nur bei höheren Wachstumsraten war der Unterschied größer, insbesondere unter anaeroben Bedingungen im engen Bereich von $0,25 \text{ h}^{-1}$ bis $0,3 \text{ h}^{-1}$.

Der ergänzende Datensatz 1 (auffindbar unter <https://github.com/Paulocracy/Dissertationsdaten> oder <https://doi.org/10.5281/zenodo.10556528>) enthält aus den Ergebnissen eine Zusammenfassung der Eigenschaften von jeweils zwei ausgewählten MDF- und SubMDF-optimalen Lösungen (je eine für hohe Wachstumsraten unter aeroben Bedingungen und eine für hohe Wachstumsraten unter anaeroben Bedingungen), zusammen mit Variabilitätsanalysen von Reaktionsflüssen und thermodynamischen Potentialen. Es zeigte sich dabei, dass typischerweise unter den gegebenen Bedingungen auftretenden Teile des Stoffwechsels, wie z.B. die Glycolyse, in den Flusslösungen tatsächlich aktiv genutzt werden und essentiell sind. Wie in diesem ergänzenden Datensatz zusätzlich gezeigt, wurden auch die möglichen Bereiche der Metabolitkonzentrationen für die MDF-optimalen Lösungen ermittelt und festgestellt, dass fast alle dieser Bereiche mit den *in vivo* gemessenen Metabolitkonzentrationen von Bennett *et al.* kompatibel sind [103].

Mit der Einzelcofaktor-Spezifität, bei der nur NAD(H) verwendet werden kann, wurden mit einer Ausnahme unter allen getesteten Bedingungen nur deutlich niedrigere MDF-Werte erreicht als mit der flexiblen und der Wildtyp-Spezifität. Schon dies deutete, im Einklang mit den Ergebnissen von [104], darauf hin, dass die Anwesenheit von zwei Redoxcofaktoren im Wildtyp zu deutlich höheren netzwerkweiten thermodynamischen Triebkräften führen kann.

Darüber hinaus zeigte der Vergleich der MDF zwischen den 1.000 zufälligen NAD(P)(H)-Spezifitäten und der Wildtyp-Spezifität, dass die Wildtyp-Spezifität thermodynamisch signifikant besser ist als zufällige Spezifitäten. In Abhängigkeit von der Wachstumsrate hatten maximal 0,1 % (aerob) bzw. 7,9 % (anaerob) der zufälligen Spezifitäten einen besseren MDF als die Wildtyp-Spezifität. Je nach Bedingung erreichten weitere zufällige Spezifitäten den gleichen MDF wie der Wildtyp, aber mindestens 83 % aller zufälligen Spezifitäten hatten einen schlechteren MDF als der Wildtyp. Interessanterweise waren unter bestimmten Bedingungen auch mehrere zufällige Spezifitäten schlechter als die Einzelcofaktor-Spezifität. Dies kann z.B. der Fall sein, wenn in einer solchen zufälligen Spezifität nur wenige und damit thermodynamisch ungünstigere Stoffwechselwege existieren, mit denen die notwendigen Mengen an NAD(H) oder NADP(H) synthetisiert werden können.

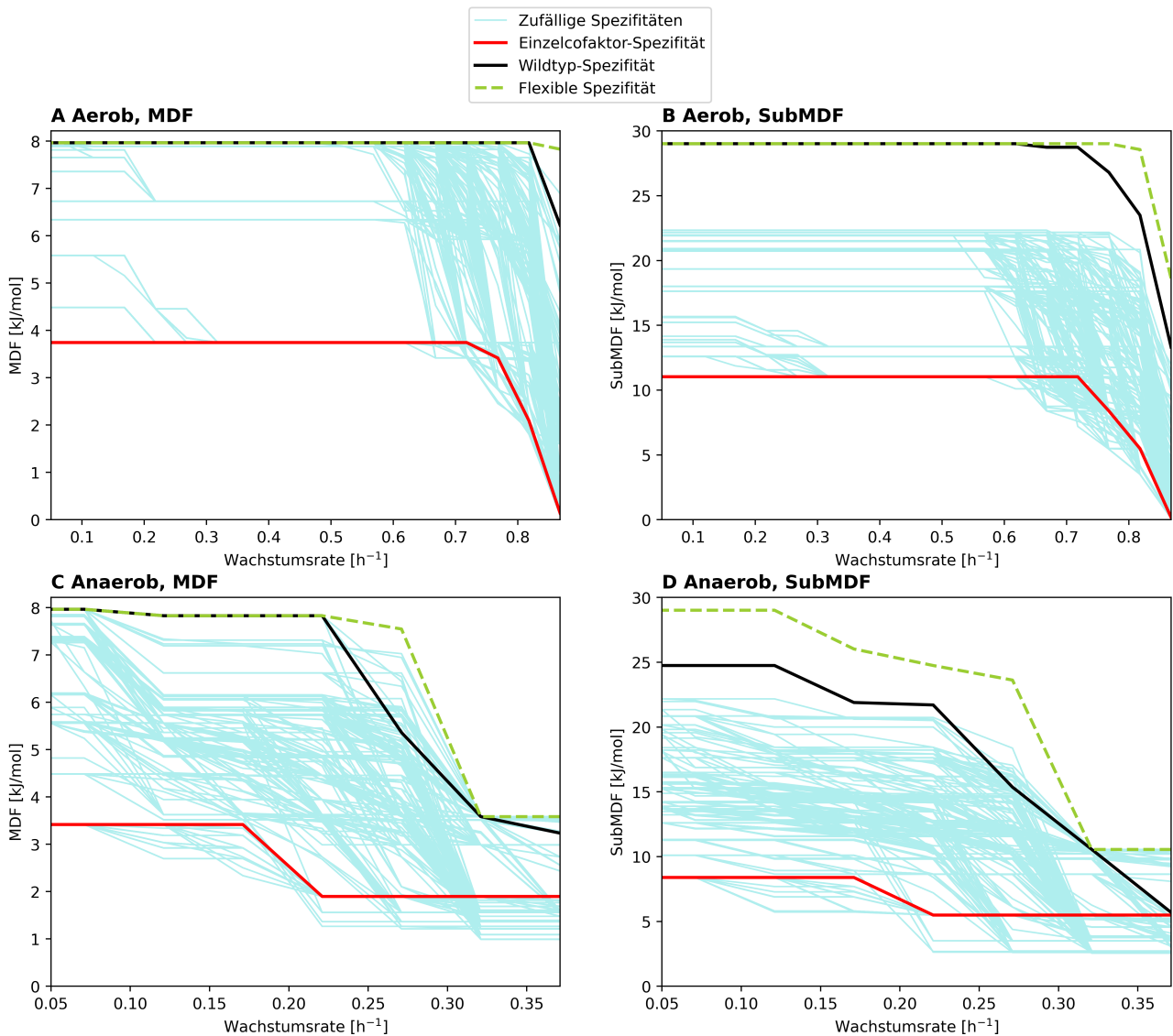


Abbildung 4.3: Maximale thermodynamische Triebkräfte (MDF und SubMDF) die mit den verschiedenen NAD(P)(H)-Spezifitäten erreichbar sind. (A) Aerobe Bedingungen mit dem MDF als Optimierungsziel; (B) Aerobe Bedingungen mit dem SubMDF als Ziel; (C) Anaerobe Bedingungen mit dem MDF als Ziel; (D) Anaerobe Bedingungen mit dem SubMDF als Ziel.

Wie bereits beschrieben (Gleichungen (4.3) bis (4.5)), wurde in diesem Kapitel zusätzlich zum MDF der SubMDF eingeführt, der den MDF für eine gewählte Teilmenge aller Reaktionen eines Netzwerks darstellt. In dieser Arbeit wurde der SubMDF für alle NAD(P)(H)-abhängigen Reaktionen des Netzwerks unter einem minimalen netzwerkweiten MDF von 0,1 kJ/mol berechnet. Damit wird der Einfluss der über das gesamte Netzwerk wirkenden thermodynamischen Flaschenhalse weitgehend umgangen, sodass man sich auf die thermodynamischen Eigenschaften der NAD(P)(H)-abhängigen Reaktionen konzentrieren kann. Wie in den Resultaten der Analyse zu erkennen ist (Abb. 4.3B sowie 5.3D; Tab. 4.2A), waren die SubMDF in der vorliegenden Analyse generell höher als die unter der jeweiligen Bedingung auftretenden MDF. Dies lag definitionsgemäß daran, dass für den SubMDF nur eine Teilmenge aller Reaktionen betrachtet wird. Obwohl die qualitativen Ergebnisse mit dem SubMDF denen mit dem MDF generell ähnlich waren, waren die absoluten Unterschiede zwischen den NAD(P)(H)-Spezifitäten deutlich ausgeprägter. So war bei

den höchsten Wachstumsraten der Unterschied zwischen der flexiblen und der Wildtyp-Spezifität im Vergleich deutlich größer. Unter aeroben Bedingungen zeigte sich, wie beim MDF, dass dieser Unterschied bei niedrigeren Wachstumsraten immer geringer wurde. Außerdem erreichte keine zufällige Spezifität bei jeglicher aerob getesteten Wachstumsrate den SubMDF der Wildtyp-Spezifität. Unter anaeroben Bedingungen hingegen lag der SubMDF der Wildtyp-Spezifität mit Ausnahme einer Wachstumsrate immer deutlich unter dem Wert der flexiblen Spezifität. Bei der höchsten Wachstumsrate erreichten 13,8 % der zufälligen Spezifitäten einen identischen oder besseren SubMDF als der Wildtyp. Bei niedrigeren Wachstumsraten erreichte hingegen keine der zufälligen Spezifitäten den Wildtypen. Grundsätzlich war bei kleinen und mittleren Wachstumsraten der relative Abstand zwischen dem SubMDF der flexiblen bzw. Wildtyp-Spezifität einerseits und den besten zufälligen Spezifitäten bzw. der Einzelcofaktor-Spezifität andererseits deutlich größer als der entsprechende Abstand beim MDF.

Tabelle 4.2: (A) Vergleich des (Sub)MDF der zufälligen Spezifitäten und der Wildtyp-Spezifität; (B) Anzahl der notwendigen Cofaktor-Austausche der Wildtyp-Spezifität um den optimalen (Sub)MDF der flexiblen Spezifität zu erreichen. Es wurden 1.000 zufällige Spezifitäten berechnet.

Sauerstoffbedingung	Wachstumsrate [h ⁻¹]	(A) Anteil der zufälligen Spezifitäten mit höherem/gleichem/niedrigerem (Sub)MDF im Vergleich zur Wildtyp-Spezifität		(B) Anzahl der notwendigen Cofaktor-Austausche im Wildtypen um den theoretisch optimalen (Sub)MDF zu erreichen	
		MDF	SubMDF	MDF	SubMDF
aerob	0,87	0,1 % / 0 % / 99,9 %	0 % / 0 % / 100 %	6	2
aerob	0,82	0 % / 0,4 % / 99,6 %	0 % / 0 % / 100 %	0	3
aerob	0,77	0 % / 2,3 % / 97,7 %	0 % / 0 % / 100 %	0	2
aerob	0,72	0 % / 5,6 % / 94,4 %	0 % / 0 % / 100 %	0	2
aerob	0,67	0 % / 11,8 % / 88,2 %	0 % / 0 % / 100 %	0	0
aerob	0,62	0 % / 14,1 % / 85,9 %	0 % / 0 % / 100 %	0	0
aerob	0,568 ... 0,518	0 % / 14,9 % / 85,1 %	0 % / 0 % / 100 %	0	0
aerob	0,468 ... 0,118	0 % / 15,0 % / 85,0 %	0 % / 0 % / 100 %	0	0
aerob	0,07	0 % / 17,0 % / 83,0 %	0 % / 0 % / 100 %	0	0
aerob	0,05	0 % / 17,0 % / 83,0 %	0 % / 0 % / 100 %	0	0
anaerob	0,37	7,9 % / 0,3 % / 92,8 %	13,4 % / 0,4 % / 86,2 %	1	1
anaerob	0,32	0 % / 8,7 % / 91,3 %	0 % / 7,1 % / 92,9 %	0	0
anaerob	0,27	3,7 % / 0,4 % / 95,9 %	1,8 % / 0,3 % / 97,9 %	9	14
anaerob	0,22	0 % / 0,6 % / 99,4 %	0 % / 0 % / 100 %	0	2
anaerob	0,17	0 % / 0,6 % / 99,4 %	0 % / 0 % / 100 %	0	5
anaerob	0,12	0 % / 0,6 % / 99,4 %	0 % / 0 % / 100 %	0	4
anaerob	0,07	0 % / 2,3 % / 97,7 %	0 % / 0 % / 100 %	0	3
anaerob	0,05	0 % / 2,3 % / 97,7 %	0 % / 0 % / 100 %	0	3

Zusammengefasst deuten die Ergebnisse in Abb. 4.3 sowie Tab. 4.2A sehr stark darauf hin, dass (i) das metabolische Netzwerk mit zwei Redoxcofaktorpaaren deutlich höhere thermodynamische Triebkräfte erzielen kann als ein Netzwerk mit nur einem Cofaktorpaar und (ii) die Wildtyp-Spezifität thermodynamische Triebkräfte ermöglicht, die entweder sehr nahe oder identisch am theoretischen Optimum liegen und zudem mit sehr hoher Wahrscheinlichkeit deutlich bessere Werte als eine zufällige Spezifität erzielen.

Wie viele Cofaktor-Austausche sind, ausgehend vom Wildtypen, notwendig, um die theoretisch maximalen Triebkräfte zu erreichen?

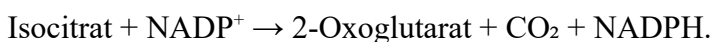
Um ein Maß für die Entfernung der Wildtyp-Spezifität von der thermodynamisch optimalen (flexiblen) Spezifität zu erhalten, wurde über ein entsprechendes MILP die Anzahl der notwendigen Cofaktor-Austausche des Wildtyps bestimmt, um einen theoretisch optimalen (Sub)MDF zu erreichen (siehe Gleichungen (4.9) bis (4.10)). Über alle getesteten Sauerstoffbedingungen und Wachstumsraten wurde in 47 der 52 resultierenden Fälle festgestellt, dass nur drei oder weniger Austausche in der Wildtyp-Spezifität dafür notwendig waren (Tab. 4.2B). Darüber hinaus war in Übereinstimmung mit Abb. 4.3 in 38 Fällen nicht einmal ein Austausch erforderlich. Eine Ausnahme bildete z.B. die Wachstumsrate $0,271 \text{ h}^{-1}$ unter Anaerobie, bei der eine größere Anzahl von Austauschen (9 bzw. 14) notwendig war, damit der Wildtyp den theoretisch optimalen MDF bzw. SubMDF erreichen konnte.

Einige spezifische Austausche traten in den Lösungen mit größerer Häufigkeit auf. Ein Beispiel ist die NAD-abhängige Pyruvatdehydrogenase (PDH) mit der Reaktionsgleichung (und der Abkürzung CoA für Coenzym A)



Auf den ersten Blick könnte dieser Austausch zu einer geringeren netzwerkweiten thermodynamischen Triebkraft führen, da, wie *in vivo*, das Verhältnis $[\text{NADH}]:[\text{NAD}^+]$ in den vorliegenden Analysen deutlich niedriger war als das von $[\text{NADPH}]:[\text{NADP}^+]$ (siehe hierzu nächsten Abschnitt). Dennoch hat dieser Austausch keinen negativen Effekt auf den (Sub)MDF. Dies kann durch den $\Delta_r G'^{\circ}$ -Wert von PDH erklärt werden, der mit $-34,47 \text{ kJ/mol}$ besonders negativ ist. Dadurch kann ein thermodynamisch ungünstiger erscheinendes Cofaktorpaar eingesetzt werden, ohne dass die PDH-Reaktion zum thermodynamischen Flaschenhals für den (Sub)MDF wird. Darüber hinaus eröffnet der Cofaktor-Austausch von PDH einen neuen Weg zur NADPH-Produktion, wodurch andere, thermodynamisch ineffizientere Wege für diese Produktion überflüssig werden. Dies wiederum kann die (Sub)MDF verbessern. Dieser Cofaktor-Austausch kann daher nicht auf der Ebene der Reaktion, sondern nur auf der Ebene des gesamten Netzwerks verstanden werden.

Ein weiteres häufiger auftauchendes Beispiel für einen Cofaktoraustausch betrifft die Reaktion der NADP-abhängigen Isocitrat-Dehydrogenase der Form



Die thermodynamische Erklärung für diesen vorgeschlagenen Austausch zu NAD(H) ergibt sich hier direkt aus den auftretenden niedrigen $[NADH]:[NAD^+]$ -Verhältnissen, durch die der positive $\Delta_r G^\circ$ von +5,12 kJ/mol dieser Reaktion überwunden werden kann. Interessanterweise sind NAD-abhängige Isocitrat-Dehydrogenasen aus anderen Organismen bekannt, die wiederum als Vorfahren der heutigen bakteriellen NADP-abhängigen Isocitrat-Dehydrogenasen angesehen werden; dennoch zeigten experimentelle Arbeiten an *E. coli*, dass ein Wechsel der Cofaktorabhängigkeit der Isocitrat-Dehydrogenase von NADP zu NAD insbesondere unter Acetat als Kohlenstoffsubstrat zu einer deutlich reduzierten Wachstumsrate führt [108]. Dies zeigt, dass einige Wildtyp-Reaktionsspezifitäten, die in dem vorliegenden Kapitel thermodynamisch nachteilig erscheinen, unter anderen, hier nicht betrachteten Wachstumsbedingungen wieder erklärbar werden könnten. Eine Möglichkeit für die Zelle, die Cofaktor-Spezifität je nach Wachstumsbedingung anzupassen, wäre die Verwendung von Enzymen, die die gleiche Reaktion katalysieren, sich aber in der NAD(P)(H)-Spezifität unterscheiden. Im *iML1515*-Modell existieren 33 solcher Reaktionspaare, von denen 22 promiskuöse Aktivitäten des jeweiligen Enzyms und nur 11 tatsächlich von jeweils unterschiedlichen Enzymen mit unterschiedlicher NAD(P)(H)-Spezifität katalysiert werden. Von diesen Reaktionspaaren existiert jedoch nur eines im zentralen Kohlenstoffwechsel, das durch die Varianten der Malatdehydrogenase SfcA (NAD-abhängig) und MaeB (NADP-abhängig) katalysiert wird und in dem jeweils Malat zu Pyruvat decarboxyliert wird [109]. Somit scheint die erwähnte Möglichkeit einer bedingungsabhängigen Anpassung der NAD(P)(H)-Spezifität zumindest im zentralen Kohlenstoffwechsel von *E. coli* letztlich von untergeordneter Bedeutung zu sein.

Effekte einzelner Cofaktoraustausche

Als weitere Analyse zur Beurteilung der Fähigkeit der Wildtyp-Spezifität im Hinblick auf das Erreichen des (Sub)MDF wurde zusätzlich eine Analyse des Cofaktoraustausches der einzelnen Reaktionen durchgeführt (Tab. 4.3). Dazu wurde für jede NAD(P)(H)-abhängige Reaktion, mit Ausnahme der Reaktionen, in denen NAD(H) und NADP(H) gleichzeitig auftreten, die NAD(P)(H)-Spezifität einzeln zum jeweils analogen Cofaktor des anderen Paares verändert. Der Effekt dieses einzelnen Austausches wurde dann durch Neuberechnung der (Sub)MDF und Vergleich mit den ursprünglichen (Sub)MDF analysiert. Jeder Austausch, der zu einer thermodynamisch unmöglichen Lösung führte, wurde aus weiteren Berechnungen ausgeschlossen.

Zur besseren Interpretation der Ergebnisse wurde aus der Menge aller in *iML1515* auftretenden NAD(P)(H)-abhängigen Reaktionen über eine Flussvariabilitätsanalyse mit thermodynamischen Nebenbedingungen, insbesondere dem Setzen eines zu erreichenden jeweiligen (Sub)MDF, die Anzahl aller unter allen betrachteten Bedingungen beteiligten (aktiven) Reaktionen ermittelt. Es zeigte sich, dass 161 dieser 238 Reaktionen zumindest zu einer (Sub)MDF-optimalen Lösung beitragen können. Als Nächstes wurde die Anzahl der Reaktionen berechnet, die unter mindestens einer der Bedingungen einen Einfluss auf den (Sub)MDF haben, wenn ihre NAD(P)(H)-Spezifität ausgetauscht wird. Insgesamt wurde ein solcher Einfluss für 71 der 161 (41 %) Reaktionen ermittelt. Tabelle 4.3 zeigt einige Kennzahlen und das Vorzeichen des Einflusses, getrennt nach Aerobie und (Sub)MDF. Wie zu sehen ist, hatten die meisten Austausche mit (Sub)MDF-Einfluss einen negativen Effekt auf den (Sub)MDF, was darauf hindeutet, dass die Wildtyp-Spezifität in Bezug auf die thermodynamische Triebkraft größtenteils vorteilhaft war. Einige einzelne

Austausche verursachten sogar drastischere negative Effekte, bis zu -5,9 kJ/mol beim MDF und sogar -18,4 kJ/mol beim SubMDF. Auch wenn einige einzelne Cofaktor-Austausche einen positiven Effekt von bis zu +1,6 kJ/mol auf den MDF und +4,9 kJ/mol auf den SubMDF hatten, reichte ein solcher Austausch in den meisten Fällen nicht aus, um den theoretisch optimalen (Sub)MDF der flexiblen Spezifität zu erreichen, wie schon im vorletzten Abschnitt gezeigt wurde (vgl. Tab. 4.2B).

Tabelle 4.3: Anzahl der Reaktionen (und deren Anteil an allen 161 NAD(P)H-abhängigen Reaktionen in iML1515, die an mindestens einer (Sub)MDF-optimale Lösung beteiligt sind), in denen ein einzelner Cofaktoraustausch den (Sub)MDF beeinflusst. Zusätzlich sind die maximalen und mittleren (Sub)MDF-Erhöhungen bzw. -Erniedrigungen angegeben.

Sauerstoffbedingung	Anzahl der Reaktionen bei denen ein Cofaktoraustausch den (Sub)MDF der Wildtyp-Spezifität beeinflusst			
	MDF		SubMDF	
	Erhöhung	Erniedrigung	Erhöhung	Erniedrigung
aerob	1 (0,6 %) +0,3 kJ/mol	16 (9,9 %) bis zu -5,9 kJ/mol Mittelwert: -2,6 kJ/mol	7 (4,3 %) bis zu +3,3 kJ/mol Mittelwert: +1,1 kJ/mol	51 (31,7 %) bis zu -18,4 kJ/mol Mittelwert: -6,5 kJ/mol
anaerob	3 (1,9 %) bis zu +1,6 kJ/mol Mittelwert: +0,6 kJ/mol	25 (15,5 %) bis zu -5,9 kJ/mol Mittelwert: -1,7 kJ/mol	9 (5,6 %) bis zu +4,9 kJ/mol Mittelwert: +1,7 kJ/mol	36 (22,4 %) bis zu -16,3 kJ/mol Mittelwert: -3,6 kJ/mol

Vorhersage qualitativer Trends der Verhältnisse von Cofaktorkonzentrationen

Im nächsten Schritt wurde untersucht, ob das Modell iML1515_TCOSA auch allgemein bekannte Trends des Oxidations- bzw. Reduktions-Levels des NAD(H)- und des NADP(H)-Pools vorhersagen kann. Da der NAD(H)-Pool in *E. coli* typischerweise deutlich stärker oxidiert ist als der NADP(H)-Pool [103], wäre zu erwarten, dass das Modell mit der Wildtyp-Spezifität für das Konzentrationsverhältnis $[NADH]:[NAD^+]$ einen niedrigeren Wert vorhersagt als für $[NADPH]:[NADP^+]$. Um diese Annahme zu bestätigen, wurden die minimal und maximal möglichen Bereiche des Quotienten der beiden Verhältnisse, d. h. $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$, unter den verschiedenen bereits getesteten Bedingungen und unter Einhaltung der entsprechenden optimalen (Sub)MDF berechnet (siehe Gleichungen (4.4) bis (4.8)). Wie in Abbildung 4.4 zu sehen ist, kann der *in vivo* bekannte Trend ($[NADH]:[NAD^+] < [NADPH]:[NADP^+]$, d.h. $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]} < 1$) tatsächlich unter allen getesteten Bedingungen mit der Wildtyp-Spezifität mit dem Modell vorhergesagt werden.

In den meisten Fällen ist das Verhältnis von reduzierten zu oxidierten Cofaktoren im NADP(H)-Pool sogar mehr als tausendfach höher als im NAD(H)-Pool, was gut mit *in vivo*-Messungen übereinstimmt (z.B. ≈ 1.500 für *E. coli* in [103]). Das höchste berechnete maximale Verhältnis beträgt $\approx 0,85$ und resultiert aus der höchsten Wachstumsrate unter anaeroben Bedingungen. Bei

diesem Wert ist zu berücksichtigen, dass die (Sub)MDF bei den höchsten Wachstumsraten in der Regel auch die niedrigsten sind (Abb. 4.3), was gleichzeitig impliziert, dass die thermodynamischen Nebenbedingungen das Reduktions- zu Oxidationsverhältnis in den Cofaktor-Pools weniger stark beeinflussen und andere thermodynamische Flaschenhälse dominieren.

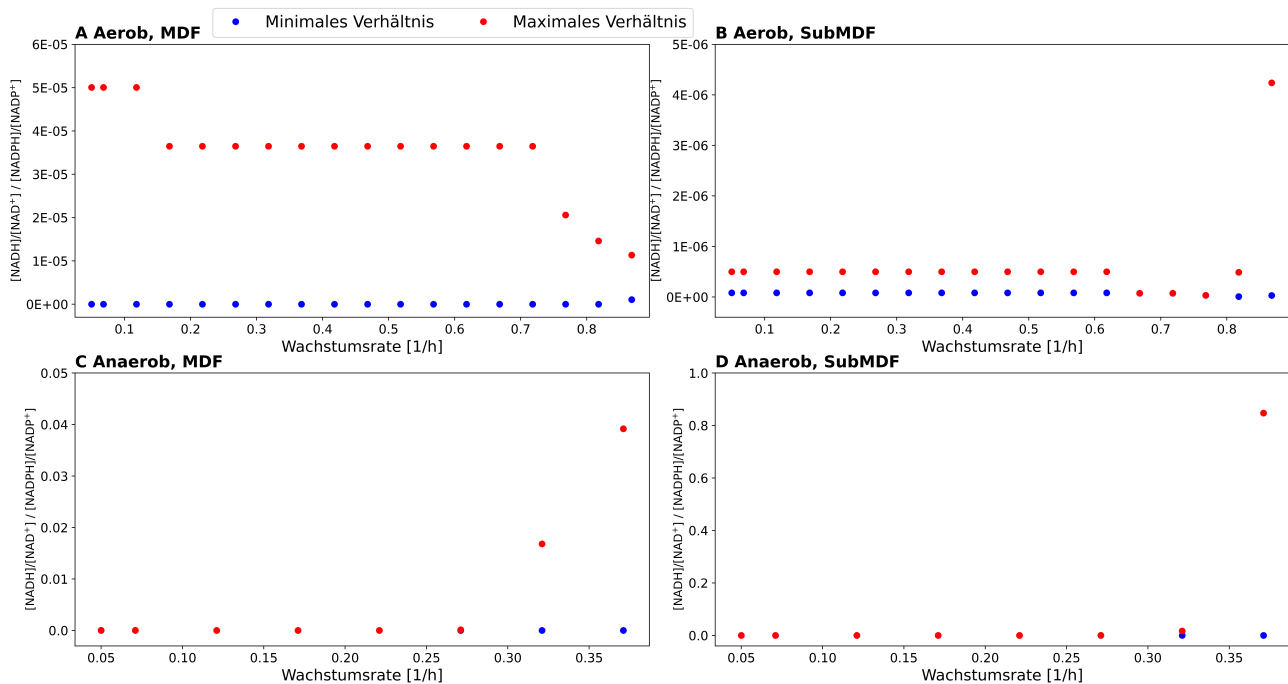


Abbildung 4.4: Minimal und maximal mögliche $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnisse mit denen der zur Wachstumsrate ermittelte (Sub)MDF mit der Wildtyp-Spezifität erreicht werden kann. (A) Aerobe Bedingungen unter MDF; (B) Aerobe Bedingungen unter SubMDF; (C) Anaerobe Bedingungen unter MDF; (D) Anaerobe Bedingungen unter SubMDF. Wann immer ein blauer Punkt (minimales Verhältnis) nicht sichtbar ist überlappt dieser weitgehend mit einem roten (maximales Verhältnis) und repräsentiert somit einen sehr ähnlichen Wert. In all diesen Fällen ist das maximale $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnis unter 0,001.

Analysen mit drei Cofaktorpaaren

Zusätzlich wurde untersucht, ob die bereits beschriebene Variante iML1515_3TCOSA, die zusätzlich zu NAD(P) ein hypothetisches drittes Redoxcofaktorpaar enthält, im Vergleich zu iML1515_TCOSA höhere thermodynamische Triebkräfte aufweisen kann. Hierzu wurde die flexible Spezifität der (Sub)MDF in iML1515_3TCOSA unter den mit iML1515_TCOSA berechneten Bedingungen bestimmt und die Ergebnisse verglichen. Ein wichtiger Aspekt bei der Einführung des dritten Cofaktorpaares ist die Wahl seines Standard-Redoxpotentials ($\Delta E'^{\circ}$). Sowohl bei NADH zu NAD⁺ als auch für NADPH zu NADP⁺ ist das $\Delta E'^{\circ}$ jeweils identisch mit einem Wert von -320 mV, was wiederum einer Differenz der Standard-Gibbs-Bildungsenergie ($\Delta_r G'^{\circ}$) von ca. 61,75 kJ/mol zwischen reduzierter und oxidierter Cofaktorvariante entspricht. In der vorliegenden Analyse wurden daraus abgeleitet die folgenden drei Szenarien bezüglich des $\Delta E'^{\circ}$ des dritten

Cofaktorpaars betrachtet: 1) ein mit NAD(P)(H) identischer Wert von -320 mV, womit das dritte Cofaktorpaar ähnliche Eigenschaften wie NAD(P)(H) aufweist, 2) ein im Vergleich zu NAD(P)(H) um 155 mV erhöhter $\Delta E'^{\circ}$ von -165 mV und 3) ein im Vergleich um 155 mV erniedrigter $\Delta E'^{\circ}$ von -475 mV. Die Fälle 2 und 3 entsprechen einer Erniedrigung bzw. Erhöhung des $\Delta_r G^{\circ}$ aller Reaktionen mit dem dritten Cofaktorpaar um 30 kJ/mol, multipliziert mit der jeweiligen reaktionsinternen Stöchiometrie der oxidierten Cofaktorvariante.

Die Ergebnisse dieser Berechnungen sind in Abbildung 4.5 dargestellt. Wie leicht zu erkennen ist, führte das zusätzliche dritte Cofaktorpaar nur zu sehr geringen thermodynamischen Vorteilen, wenn der gleiche $\Delta E'^{\circ}$ wie für NAD(P)(H) verwendet wurde. In diesem Falle konnte kein einziger MDF-Vorteil gegenüber dem Modell mit zwei Cofaktorpaaren gefunden werden, und dazu nur kleine SubMDF-Vorteile bei wenigen (aeroben) oder sogar nur einer (anaeroben) Wachstumsrate. Ein ähnliches Ergebnis ergab sich auch bei einem erhöhten $\Delta E'^{\circ}$ von -165 mV für das dritte Cofaktorpaar. Auch hier konnte nur ein kleiner Vorteil für SubMDF ermittelt werden, diesmal sogar nur bei wenigen Wachstumsraten unter aeroben Bedingungen.

Ein deutlich anderes Bild ergab sich, wenn das $\Delta E'^{\circ}$ des dritten Cofaktorpaars auf -475 mV abgesenkt wurde. In diesem Falle war der SubMDF bei vielen Wachstumsraten signifikant erhöht, um bis zu ca. 15 kJ/mol aerob und 7 kJ/mol anaerob. Darüber hinaus konnten einige kleinere MDF-Vorteile sowohl unter aeroben als auch unter anaeroben Bedingungen festgestellt werden.

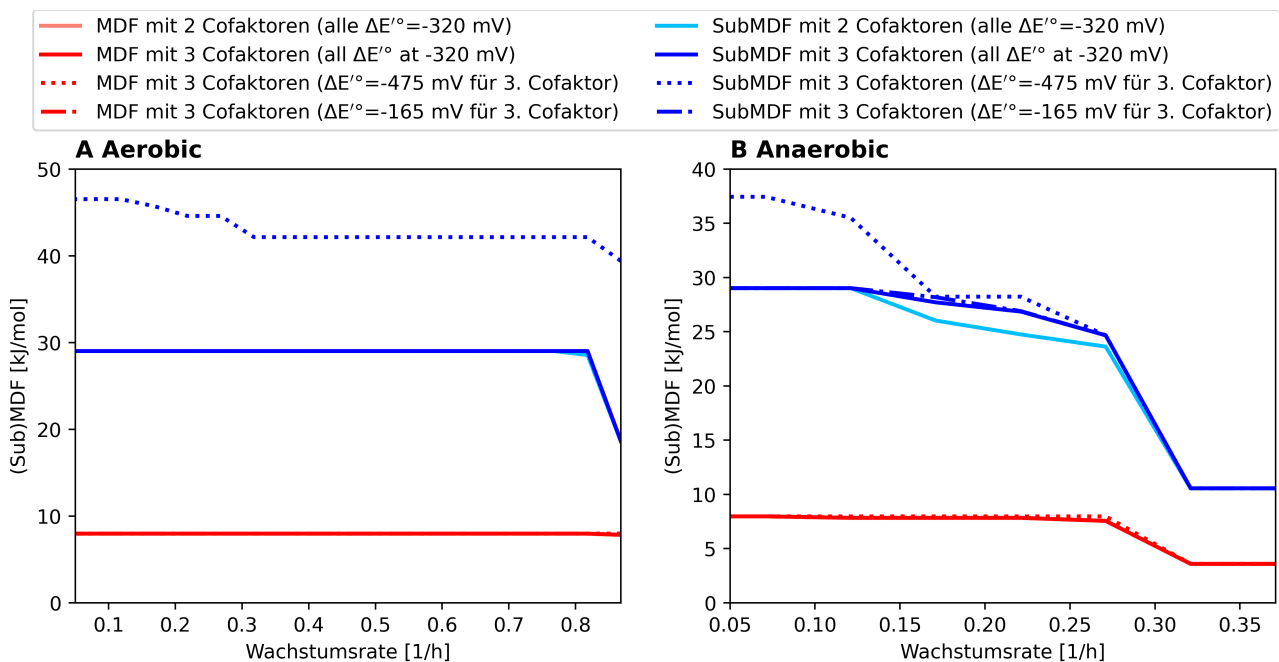


Abbildung 4.5: Vergleich des Modells mit zwei Redoxcofaktoren mit dem Modell mit drei Cofaktoren hinsichtlich der erreichbaren (Sub)MDF-Werte mit dem flexiblen Szenario. Für das Modell mit drei Cofaktoren wurden verschiedene Szenarien bezüglich des Redoxpotenzials betrachtet (siehe Text). (A) Aerobe Bedingungen; (B) Anaerobe Bedingungen. Häufig fallen die roten Linien (MDF) mit der dunkelroten Linie zusammen, ebenso wie einige blaue Linien (SubMDF) mit der dunkelblauen Linie.

Analysen hinsichtlich der Robustheit der Ergebnisse mit Glucose

Obwohl die bisher in diesem Kapitel vorgestellten Berechnungen nur wenige Parameter erfordern, nämlich die Netzwerkstöchiometrie, die Standard-Gibbs-Energien der Reaktionen und vorgegebene Bereiche von Metabolitkonzentrationen, können die Berechnungen dennoch durch Unsicherheiten beeinflusst werden. Entsprechende Robustheitsanalysen wurden daher, wie in den nächsten Unterabschnitten dargelegt, für die genutzten $\Delta_r G'^{\circ}$ sowie Metabolitkonzentrationsbereiche durchgeführt.

Ergebnisse mit variierten Standard-Gibbs-Energien

Eine wesentliche Quelle der Unsicherheit der hier dargestellten Ergebnisse betrifft die mit der eQuilibrator-API [45] ermittelten $\Delta_r G'^{\circ}$ -Werte. Darüber hinaus mussten, wie bereits erwähnt, Reaktionen ohne ermittelten $\Delta_r G'^{\circ}$ gesondert behandelt werden. Zudem können komplexere Möglichkeiten zur Umgehung thermodynamischer Flaschenhälse, wie z.B. die typischerweise große Menge an direkt an Enzyme gebundenem NAD(P)(H) [110] oder das „Kanalieren“ (englisch *channeling*) von Metaboliten über mehrere Reaktionen hinweg [111], in den hier durchgeführten Analysen nicht berücksichtigt werden. Um den Effekt veränderter $\Delta_r G'^{\circ}$ auf die Ergebnisse abschätzen zu können, wurde eine entsprechende Robustheitsanalyse durchgeführt.

Diese Robustheitsanalyse mit zufällig variierten $\Delta_r G'^{\circ}$ -Werten wurde wie folgt durchgeführt: Um die thermodynamische Konsistenz auch bei variierten $\Delta_r G'^{\circ}$ -Werten beizubehalten, wurde unter Verwendung der grundlegenden Überlegungen dahingehend in [45], ein Zufallswert zu $\Delta_f G'^{\circ}$, der Standard-Gibbs-Formationsenergie eines Metaboliten, addiert. Die Beziehung zwischen $\Delta_r G'^{\circ}$ und $\Delta_f G'^{\circ}$ (im Folgenden als Vektoren aller Gibbs-Energien dargestellt) lautet dabei

$$\Delta_r G'^{\circ} = N^T \cdot \Delta_f G'^{\circ}.$$

Dementsprechend können sogar ohne Kenntnis der $\Delta_f G'^{\circ}$ die $\Delta_r G'^{\circ}$ in einer thermodynamisch konsistenten Weise variiert werden, hier notiert als $\Delta_r \hat{G}'^{\circ}$. Hierzu wird ein zufälliger Wert δ zu den $\Delta_f G'^{\circ}$ hinzugefügt:

$$\Delta_r \hat{G}'^{\circ} = N^T \cdot (\Delta_f G'^{\circ} + \delta) = N^T \cdot \Delta_f G'^{\circ} + N^T \cdot \delta = \Delta_r G'^{\circ} + N^T \cdot \delta.$$

Die Werte in δ wurden einer uniformen Zufallsverteilung im Bereich von -5 kJ/mol bis +5 kJ/mol entnommen. Die resultierende Variation des $\Delta_r G'^{\circ}$ einer Reaktion kann entsprechend höher sein, z.B. für eine Reaktion $A + B \rightarrow C + D$ in einem Bereich bis zu ± 20 kJ/mol. Als einzige Ausnahme wurden die $\Delta_f G'^{\circ}$ der ubiquitär vorkommenden kleinen Moleküle H^+ , H_2O , O_2 sowie CO_2 nicht variiert. Auf die beschriebene Weise wurden insgesamt 100 variierte Mengen an $\Delta_r G'^{\circ}$ erzeugt. Für jede dieser Mengen wurden die (Sub)MDF-Werte mit Einzelcofaktor-, Wildtyp- und flexibler Spezifität berechnet.

Abb. 4.6 zeigt die Ergebnisse mit den variierten $\Delta_r G'^{\circ}$ -Werten. Analog zu Abb. 4.3 ohne variierte $\Delta_r G'^{\circ}$ war deutlich zu erkennen, dass unter allen getesteten Bedingungen der mittlere (Sub)MDF für

die Wildtyp-Spezifität immer nahe oder sogar identisch mit dem mittleren (Sub)MDF der flexiblen Spezifität und immer deutlich höher als der mittlere (Sub)MDF für die Einzelcofaktor-Spezifität war. Interessanterweise war bei allen Spezifitäten der mittlere (Sub)MDF fast immer niedriger als der entsprechende Wert ohne Variation der $\Delta_r G^\circ$. Dies könnte auf die hohe Wahrscheinlichkeit zurückzuführen sein, dass eine Erhöhung des $\Delta_r G^\circ$ einer essentiellen Reaktion zu einem thermodynamischen Flaschenhals führen konnte. Dagegen erschien es weniger wahrscheinlich, dass der $\Delta_r G^\circ$ -Wert aller thermodynamischen Flaschenhalse verringert wird.

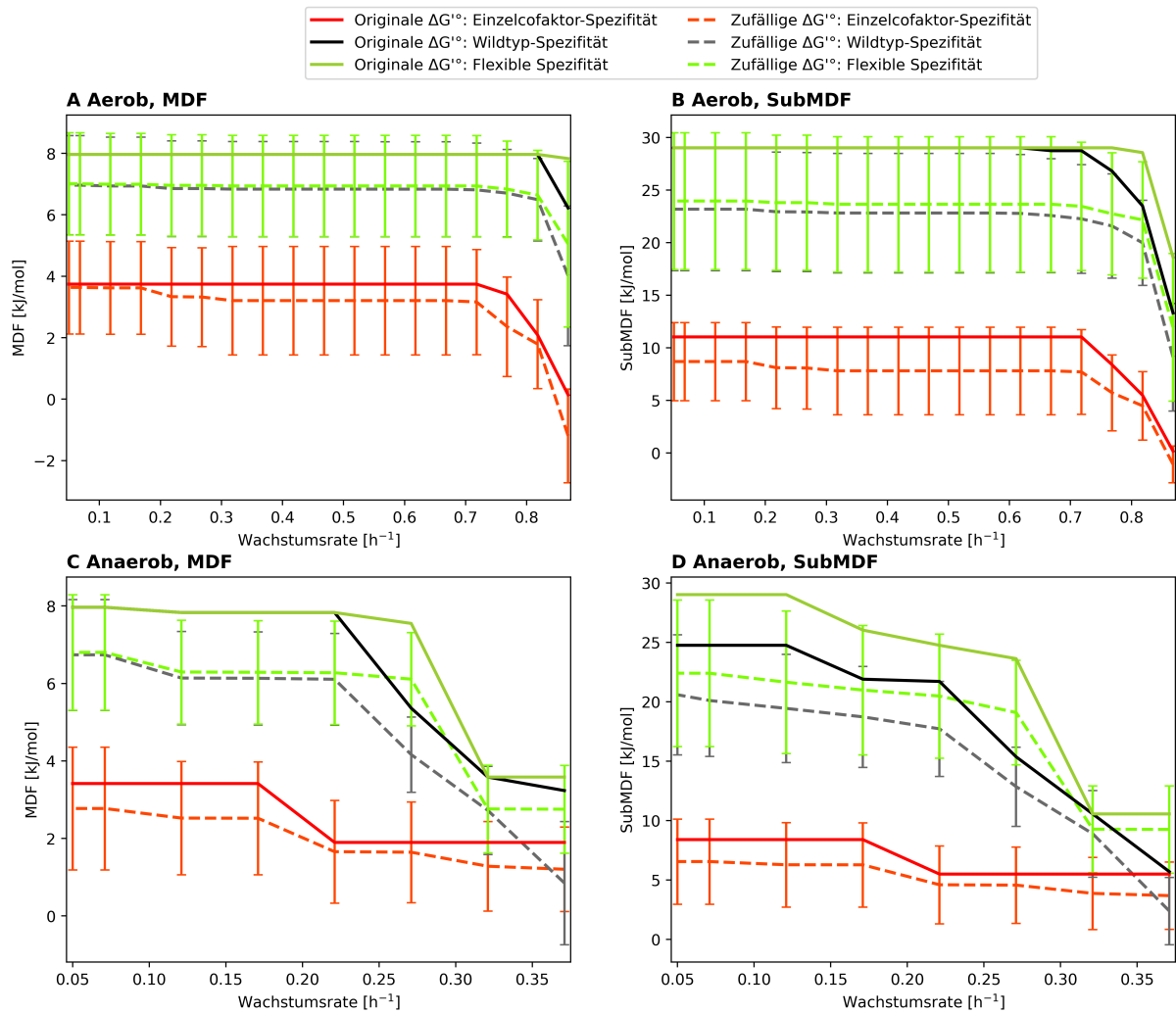


Abbildung 4.6: Vergleich der (Sub)MDF-Resultate für 100 randomisiert verrauschte Mengen an $\Delta_r G^\circ$ -Werten (gestrichelte Linie) mit denen mit den originalen $\Delta_r G^\circ$ -Werten aus Abb. 4.3. (A) Aerobe Bedingungen mit dem MDF als Optimierungsziel; (B) Aerobe Bedingungen mit dem SubMDF als Ziel; (C) Anaerobe Bedingungen mit dem MDF als Ziel; (D) Anaerobe Bedingungen mit dem SubMDF als Ziel. Die Fehlerbalken zeigen die Standardabweichung der Resultate mit allen 100 getesteten randomisiert verrauschten Mengen an $\Delta_r G^\circ$ -Werten.

Ergebnisse mit *in vivo* gemessenen Metabolitkonzentrationen

Um die Sensitivität der Ergebnisse in Bezug auf die gewählten Bereiche der Metabolitkonzentrationen zu analysieren, wurden die wichtigsten Berechnungen dieses Kapitels auch mit (in der Regel engeren) *in vivo* gemessenen Metabolitkonzentrationen durchgeführt. Als Grundlage diente eine entsprechende Messung in *E. coli* unter aeroben Bedingungen und mit Glucose als Kohlenstoffsubstrat [103]. Das dort angegebene 95 %-Konfidenzintervall der Konzentrationsmessungen diente als neue untere und obere Schranke für jeden Metaboliten. NAD(P)(H) sowie alle Metabolite, deren Konzentrationen in der Messung nicht bestimmt wurden, erhielten die bereits vorher genutzten Standardkonzentrationsbereiche, und alle anderen Parameter blieben unverändert.

Genauso wie mit den Standardkonzentrationen mussten die $\Delta_r G^{\circ}$ einiger Reaktionen gelockert werden, um Wachstum thermodynamisch zu ermöglichen. Die Menge der auf die gleiche Weise wie mit den Standardkonzentrationen ermittelten thermodynamischen Flaschenhalse enthält tatsächlich alle 9 mit Standardkonzentrationen ermittelten Reaktionen (Tab. 4.1) plus drei zusätzliche Reaktionen (Tab. 4.4).

Tabelle 4.4: Liste der zusätzlich zu den in Tab. 4.1 in ihrem $\Delta_r G^{\circ}$ gelockerten Reaktionen, durch die im Modell Wachstum unter gemessenen Metabolitkonzentrationen aus [103] thermodynamisch möglich wird. Der $\Delta_r G^{\circ}$ von ASAD sowie GLUDy wurde auf +100 kJ/mol gesetzt da jeweils deren Rückwärtsrichtung einen thermodynamischen Flaschenhals darstellt. Alle drei hier angegebenen Reaktionen sind reversibel.

BiGG-ID der Reaktion	Name des Enzyms der Reaktion	Durch eQuilibrator ermittelter $\Delta_r G^{\circ}$ [kJ/mol]	Korrigierter $\Delta_r G^{\circ}$ [kJ/mol]
ASAD	Aspartat-Semialdehyd-Dehydrogenase	-24,32	100
GLUDy	Glutamat-Dehydrogenase (NADP)	+1809,64	100
CBMKr	Carbamat-Kinase	+278,17	-100

Wie bei den Standardkonzentrationsbereichen wurden auch bei den gemessenen Konzentrationsbereichen die (Sub)MDF über alle getesteten Spezifitäten bestimmt. Für die MDF zeigte sich, dass aufgrund der grundsätzlich engeren Konzentrationsbereiche ein einziger starker thermodynamischer Flaschenhals (der nicht in den Redoxreaktionen liegt) den MDF bei fast allen Spezifitäten und Wachstumsraten auf einen relativ niedrigen Wert von 1,9 kJ/mol begrenzt. Daher konzentrierte sich die weitere Analyse auf den SubMDF. Für diesen zeigte sich, dass die wesentlichen Ergebnisse mit den Standardkonzentrationsbereichen auch mit den gemessenen Konzentrationsbereichen reproduziert werden können (Abb. 4.7A). So erreichen zwar einige wenige der zufälligen Spezifitäten bei der höchsten Wachstumsrate den SubMDF der Wildtyp-Spezifität, bei allen niedrigeren Wachstumsraten ist dies jedoch bei keiner der zufälligen Spezifitäten der Fall. Im Gegensatz zum entsprechenden Fall mit Standardkonzentrationen erreicht der SubMDF der Wildtyp-Spezifität bei niedrigeren Wachstumsraten jedoch nicht ganz den theoretisch optimalen SubMDF der flexiblen Spezifität. Dies könnte auf die spezifischen

Bedingungen bei der Messung der verwendeten Metabolitkonzentrationen, wie etwa die eingestellte Wachstumsrate, zurückzuführen sein.

Auch die Berechnung des $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnisses, unter Einhaltung des jeweiligen SubMDF, erfolgte mit den gemessenen Konzentrationen wobei ebenfalls alle wesentlichen Ergebnisse mit Standardkonzentrationen ihre Gültigkeit behielten (Abb. 4.7B). So ergab sich wieder, dass der NAD(H)-Pool in allen Fällen deutlich oxidiert sein muss als der NADP(H)-Pool (zumindest um das Tausendfache). Konkret musste das $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnis in allen getesteten Fällen sehr geringe Werte von unter 0,0006 erreichen. In dem hier verwendeten *in vivo*-Messdatensatz beträgt der entsprechende Messwert für dieses Verhältnis 0,00056 [103] und liegt damit sehr nahe am berechneten Maximum.

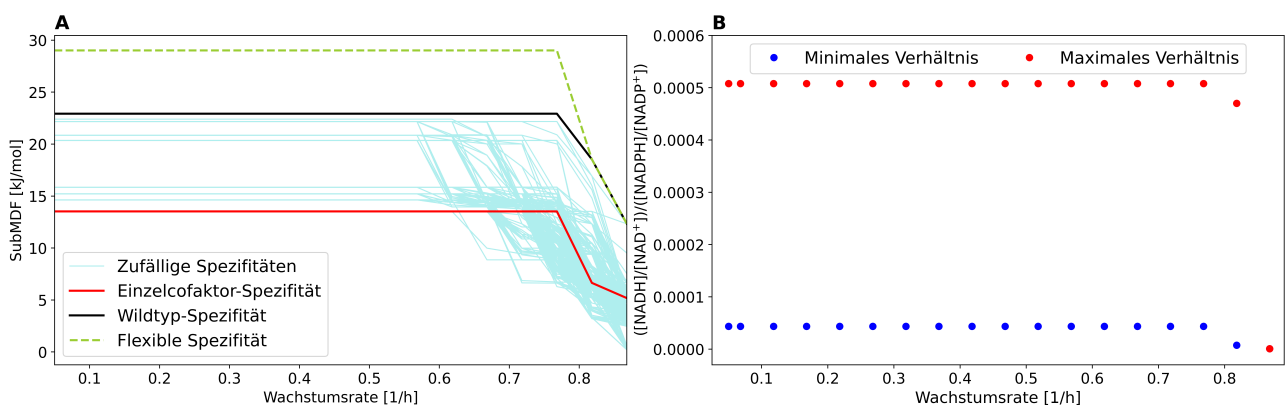


Abbildung 4.7: Berechnung der zentralen Ergebnisse unter gemessenen Metabolitkonzentrationen aus [103] und aeroben Bedingungen. (A) SubMDF-Werte verschiedene NAD(P)(H)-Spezifitäten; (B) Minimale sowie maximale $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnisse unter dem entsprechenden SubMDF.

Ergebnisse mit Acetat als Substrat

Bisher wurden alle Berechnungen mit Glucose als Kohlenstoffsubstrat durchgeführt. Um festzustellen, ob die wichtigsten Ergebnisse auch mit einer anderen Quelle reproduziert werden können, wurden einige Berechnungen mit Acetat als Kohlenstoffsubstrat wiederholt. Alle anderen Einstellungen blieben, im Vergleich zu den Berechnungen mit Glucose, unverändert. Acetat betritt an einer weiter entfernten Stelle den zentralen Kohlenstoffwechsel, wodurch für Wachstum, anstelle der Glycolyse, die Gluconeogenese benötigt wird.

Die so resultierenden (Sub)MDF-Werte unter verschiedenen NAD(P)(H)-Spezifitäten (Abb. 4.8A und 4.8C) und die möglichen Bereiche des $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnisses (Abb. 4.8B und 4.8D) führen alle zu ähnlichen grundsätzlichen Erkenntnissen wie mit Glucose als Kohlenstoff-Substrat. Konkret erreicht die Wildtyp-Spezifität wiederum unter fast allen betrachteten Bedingungen einen höheren (Sub)MDF als alle Zufallsspezifitäten. Die einzige Ausnahme ist die

maximale Wachstumsrate, bei der nur 25 (für SubMDF) oder sogar nur eine (MDF) der 1.000 Zufallsspezifitäten eine höhere (Sub)MDF erreichen. Die Betrachtung des $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnisses ergibt erneut die erwartbaren Werte deutlich unter 1.

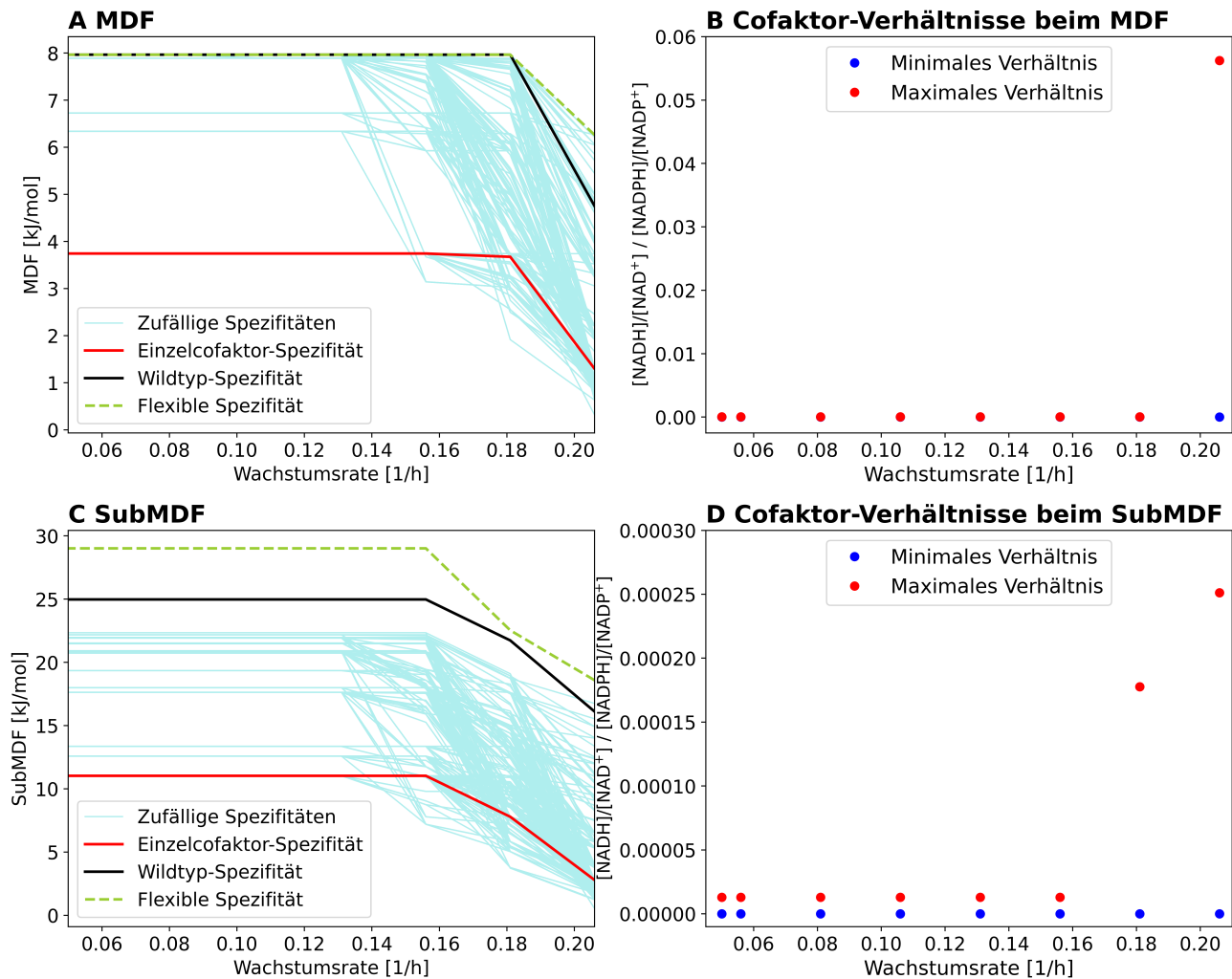
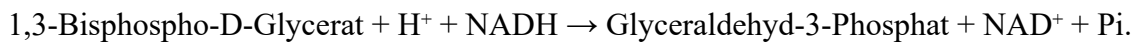


Abbildung 4.8: Maximale (Sub)MDF mit verschiedenen NAD(P)(H)-Spezifitäten sowie mögliche Bereiche des $\frac{[NADH]:[NAD^+]}{[NADPH]:[NADP^+]}$ -Verhältnisses um den jeweiligen (Sub)MDF zu erreichen mit der Wildtyp-Spezifität, alles mit Acetat als Kohlenstoff-Substrat unter aeroben Bedingungen. (A) MDF-Resultate; (B) Resultate des Verhältnisses unter dem MDF als Grenze; (C) SubMDF-Resultate; (D) Verhältnisse unter dem SubMDF. Wann immer ein blauer Punkt (minimales Verhältnis) nicht sichtbar ist überlappt dieser mit einem roten (maximales Verhältnis) und repräsentiert somit einen ähnlichen Wert.

Zusätzlich wurde wieder die minimale Anzahl von Cofaktor-Austauschen berechnet, mit der die Wildtyp-Spezifität das theoretische (Sub)MDF-Optimum der flexiblen Spezifität erreichen kann. Es zeigte sich, dass die Wildtyp-Spezifität maximal nur einen entsprechenden Austausch vom theoretischen Optimum entfernt ist. In vielen Fällen handelte es sich dabei um die umgekehrte

(gluconeogene) Richtung der Glycerinaldehyd-3-Phosphat-Dehydrogenase-Reaktion (GAPDH) der Form



Dieser Austausch zu NADP(H) kann die thermodynamische Triebkraft dieser Reaktion drastisch erhöhen, da das Konzentrationsverhältnis des resultierenden Edukts NADPH zum Produkt NADP⁺ deutlich höher ist als das von NADH zu NAD⁺ (siehe auch Abb. 4.7B und 4.7D). Dies wiederum kann den (Sub)MDF des gesamten Netzwerks verbessern. Dieses Ergebnis ist auch konsistent mit der Beobachtung, dass in vielen Organismen, in denen GAPDH typischerweise in dieser umgekehrten Richtung abläuft, wie z.B. im Calvin-Zyklus photosynthetischer Organismen, NADPH als Cofaktor bevorzugt wird. Es wäre daher interessant, *in vivo* einen entsprechenden Cofaktoraustausch für GAPDH in *E. coli* unter Acetat zu untersuchen.

Die Auswirkungen der unter Glucose gefundenen vorteilhaften Cofaktor-Austausche, für die Reaktionen PDH und Isocitratdehydrogenase, wurden auch unter Acetat untersucht. Im Falle der PDH konnte jetzt nur noch ein SubMDF-Vorteil in einer spezifischen Wachstumsrate ermittelt werden. Bei der Isocitratdehydrogenase hingegen konnte kein Vorteil und bei einigen Wachstumsraten sogar ein negativer (Sub)MDF-Effekt festgestellt werden. Insbesondere das letzte Beispiel zeigt erneut, dass für eine vollständige Beurteilung der Wildtyp-Spezifität alle relevanten Wachstumsbedingungen berücksichtigt werden müssen.

4.4 Diskussion

In diesem Kapitel wurde das TCOSA-Framework entwickelt und eingesetzt, um Schlüsselaspekte der Redoxcofaktor-Redundanz in metabolischen Netzwerken zu untersuchen. TCOSA verwendet einen unvoreingenommenen Ansatz, der ausschließlich auf einem stöchiometrischen Netzwerk, thermodynamischen Nebenbedingungen und Konzentrationsbereichen von Metaboliten basiert. Ziel der hier dargelegten Anwendung von TCOSA ist es, Beziehungen zwischen der Verteilung von NAD(P)(H)-Spezifitäten einzelner Reaktionen und netzwerkweiten thermodynamischen Triebkräften herzustellen. Die wichtigsten Ergebnisse, die aus einem genomskaligen Stoffwechselmodell von *E. coli* unter Berücksichtigung unterschiedlicher Wachstumsraten unter aeroben und anaeroben Bedingungen abgeleitet wurden, lassen sich wie folgt zusammenfassen: (1) Die Verwendung von zwei redundanten Redoxcofaktor-Pools ist offensichtlich vorteilhaft für die Maximierung der netzwerkweiten thermodynamischen Triebkraft (MDF) im gesamten Netzwerk. Zugleich ist auch der SubMDF, hier die maximierte thermodynamische Triebkraft aller NAD(P)(H)-abhängigen Reaktionen, durch die Nutzung zweiter Cofaktor-Pools deutlich erhöht. Die Einführung eines dritten Redoxcofaktors würde jedoch nur dann signifikante (Sub)MDF-Vorteile bringen, wenn dieser ein niedrigeres Redoxpotential als NAD(P)H aufweist. (2) Mithilfe von TCOSA wird gezeigt, dass die *in vivo* auftretenden NAD(P)(H)-Coenzym-spezifitäten metabolischer Reaktionen in *E. coli* wahrscheinlich das Ergebnis eines evolutionären Strebens nach einer Maximierung der netzwerkweiten thermodynamischen Triebkräfte sind. Dies wird dadurch deutlich, dass der, der mit Wildtyp-Spezifitäten erreicht werden kann, oft nahe am theoretischen Optimum liegt oder dieses sogar erreicht. (3) TCOSA ermöglicht die Vorhersage von Trends in den MDF-optimalen Verhältnissen von Redoxcofaktor-Konzentrationen, die den *in vivo*-Bedingungen

entsprechen. (4) Die erzielten Ergebnisse bleiben auch bei Variation der Modellparameter, wie z.B. der Standard-Gibbs-Energien und der Konzentrationsbereiche der Metaboliten, stabil.

Das vielleicht wichtigste Resultat der hier gezeigten Anwendung von TCOSA ist, dass die Wildtyp-Verteilung der NAD(P)(H)-Spezifitäten in den Redoxreaktionen von *E. coli* insgesamt höhere netzwerkweite thermodynamische Triebkräfte ermöglicht als fast alle zufällig erzeugten Spezifitäten (Abb. 4.3). Die Wildtyp-Spezifitäten erreichen in den meisten Fällen sogar die maximalen (Sub)MDF-Werte, die mit einer (hypothetischen) vollständig flexiblen Spezifität erreichbar sind. Dieses Ergebnis ist bemerkenswert und deutet darauf hin, dass Redoxcofaktor-Spezifitäten wahrscheinlich auf einem evolutionären Optimierungsprinzip beruhen, das die netzwerkweite thermodynamische Triebkraft maximiert. Der evolutionäre Vorteil höherer (Sub)MDF-Werte liegt dabei in der Verringerung der Wahrscheinlichkeit kleiner thermodynamischer Triebkräfte, die den Fluss bestimmter wichtiger Reaktionen möglicherweise einschränken könnten, u.a. da durch höhere (Sub)MDF auch ein breiterer Bereich an Metabolitkonzentrationen weiterhin zu thermodynamisch möglichem Wachstum führen kann. Die Bedeutung der thermodynamischen Triebkraft für den Reaktionsfluss wird deutlich, wenn man die schon beschriebene zerlegte Michaelis-Menten-Kinetik betrachtet (Gleichung (2.4)). Dementsprechend erhöhen Reaktionen, die näher am thermodynamischen Gleichgewicht ablaufen, die Anforderungen an die Menge an Enzymen, die zum Erreichen erforderlicher Flussraten erforderlich sind [104]. Da hohe MDF-Werte potentiell zu erhöhten thermodynamischen Triebkräften für alle Reaktionen führen können, wird auch der aufgrund thermodynamischer Beschränkungen erhöhte Enzymbedarf unwahrscheinlicher. Neben den thermodynamischen Effekten kann die NAD(P)(H)-Verteilung auch die möglichen stöchiometrischen Gleichgewichte der NAD(H)- und NADP(H)-erzeugenden und -verbrauchenden Reaktionen beeinflussen. Eine Änderung dieser Verteilung kann die Aktivierung neuer Wege mit potentiell niedrigeren oder höheren thermodynamischen Triebkräften erforderlich machen. Der TCOSA-Ansatz berücksichtigt solche Effekte dabei direkt in seiner Formulierung.

Dass die Wildtyp-Spezifität nicht immer den theoretisch maximalen (Sub)MDF erreicht, widerspricht nicht notwendigerweise den obigen Schlussfolgerungen. Hierbei ist zunächst wieder wichtig zu betrachten, dass sich im Szenario der flexiblen Spezifität, bei dem die maximalen (Sub)MDF-Werte erreicht werden, die Spezifitäten unter verschiedenen Wachstumsraten und Bedingungen ändern können. Dies ist innerhalb der festen Wildtyp-Verteilung nicht möglich. Da sich in den optimalen Verteilungen die flexible Spezifität unter verschiedenen Bedingungen zum Erreichen maximaler (Sub)MDF-Werte auch tatsächlich ändert, muss eine festgesetzte Spezifität schlechter abschneiden als die flexible Spezifität. Darüber hinaus könnten andere Einschränkungen, die über den modellierten Rahmen von TCOSA hinausgehen, die optimale NAD(P)(H)-Spezifität und die erreichbaren MDF beeinflussen. So könnten beispielsweise bestimmte Metabolitkonzentrationen, die für eine MDF-optimale Flussverteilung erforderlich sind, übermäßig hoch und damit potentiell toxisch sein, was den möglichen Bereich der Stoffwechselkonzentrationen stärker einschränken könnte als hier angenommen. Zudem werden die möglichen Bereiche der Metabolitkonzentrationen wesentlich durch die kinetischen Eigenschaften der Reaktionsmechanismen und Enzyme bestimmt, die auch die allosterische (und genetische) Regulation der Enzyme umfassen. Aufgrund dieser Faktoren erscheint es auch plausibel, dass die berechnete MDF der Wildtyp-

Spezifität *in vivo* nicht erreicht wird. Nichtsdestotrotz ist die Analyse von MDF-optimalen Lösungen insofern von biologischer Bedeutung, als die berechnete MDF die thermodynamische Flexibilität des Systems quantifiziert: Ein hoher MDF-Wert zeigt an, dass Metabolit-Konzentrationen aus einem breiteren Bereich ausgewählt werden können, ohne dass eine aktive Reaktion den thermodynamisch möglichen Rahmen verlässt. Dies gilt selbst dann, wenn der theoretisch machbare MDF selbst unerreicht bleibt. Trotz der oben genannten potentiellen Einschränkungen hat sich der MDF-basierte Modellierungsrahmen als äußerst wertvoll erwiesen, um verschiedene Facetten der Cofaktor-Spezifität zu untersuchen.

Ein weiteres wichtiges Ergebnis war, dass zum Erreichen der theoretisch optimalen (Sub)MDF der flexiblen Spezifität nur eine relativ kleine Anzahl von Reaktionen innerhalb der Wildtyp-Spezifität einen Cofaktoraustausch erfordern würde (Tab. 4.2B). Diese Analyse stellt beispielsweise einen Cofaktoraustausch für die Isocitrat-Dehydrogenase als vorteilhaft dar, der von NADP(H) zu NAD(H) führen würde. Ein solcher Cofaktoraustausch deckt sich mit dem, was für dieses Enzym in anderen Organismen bereits bekannt ist. Für Wachstum auf Acetat wiederum konnte jedoch kein (Sub)MDF-Vorteil unter Acetat-Aufnahme festgestellt werden. Dies deutet einmal mehr darauf hin, dass andere Faktoren zu Nachteilen führen, die in dieser Analyse nicht berücksichtigt wurden, zur Entwicklung der in *E. coli* vorhandenen Cofaktor-Spezifitäten beitragen könnten.

In den meisten Fällen zeigte die Einzelaustausche-Analyse entweder eine negative oder keine Auswirkung auf den (Sub)MDF. Die Tatsache, dass mehrere einzelne Austausche keine Auswirkung hatten, zeigt an, dass andere Cofaktor-Spezifitätsverteilungen genauso gut funktionieren könnten wie die der Wildtyp-Spezifität. Während die Wildtyp-Spezifität also weitgehend durch thermodynamische Vorteile erklärt werden kann, ist ihre genaue Vorhersage aufgrund der Existenz thermodynamisch identisch erscheinender alternativer Spezifitäten im Rahmen von TCOSA nicht möglich. Die Wahrscheinlichkeit, zufällig auf eine Spezifität zu stoßen, die zu den hohen (Sub)MDF-Werten des Wildtyps führt, ist jedoch äußerst gering, wie die Analysen mit den zufälligen Spezifitäten zeigen (Abb. 4.3). Auch lässt sich spekulieren, dass viele dieser alternativen Spezifitäten unter bestimmten Bedingungen, auf die der TCOSA-Ansatz nicht eingeht, biologische Nachteile mit sich bringen würden.

TCOSA wurde auch verwendet, um zusätzliche Aspekte im Zusammenhang mit redundanten Redoxcofaktor-Pools zu untersuchen. So konnte gezeigt werden, dass mit TCOSA qualitative Trends der Redoxcofaktor-Verhältnisse vorhergesagt werden können, insbesondere, dass der NAD(H)-Pool in der Regel deutlich stärker oxidiert ist als der NADP(H)-Pool (Abb. 4.4). Der TCOSA-Ansatz ermöglicht auch eine Analyse des Effekts anderer Anzahlen (weniger oder mehr als 2) an Cofaktoraustauschen. Die Ergebnisse dieser Analyse zeigten deutlich, dass die Verwendung eines einzigen Cofaktor-Pools die thermodynamischen Triebkräfte innerhalb des Netzwerks deutlich reduzieren würde (Abb. 4.3). Diese Erkenntnis unterstreicht den bereits postulierten Vorteil zweier Cofaktor-Pools [104], der es der Zelle ermöglicht, einen überwiegend oxidierten und einen überwiegend reduzierten Cofaktor-Pool zu besitzen. Folglich kann die Spezifität jeder Reaktion so angepasst werden, dass die netzwerkweiten thermodynamischen Antriebskräfte optimiert werden. Die Untersuchung des Modells mit drei Cofaktor-Pools, in dem neben NAD(H) und NADP(H) ein hypothetisches drittes Cofaktorpaar eingeführt wurde, zeigte, dass ein daraus resultierender

signifikanter Vorteil für die (Sub)MDF nur dann auftritt, wenn das Redoxpotential des dritten Cofaktorpaares deutlich niedriger ist als das von NAD(P)(H). Diese Beobachtung könnte auf die thermodynamische Bedeutung anderer Redoxcofaktoren in *E. coli*, wie z.B. der Thioredoxine, hinweisen. Thioredoxine sind jedoch an wesentlich weniger Stoffwechselreaktionen beteiligt als NAD(P)(H), und man könnte vermuten, dass ihre Bedeutung als Elektronenträger in erster Linie in der Erleichterung spezieller Elektronentransferreaktionen liegt, wie z.B. der Reduktion von Disulfidbrücken in Proteinen. Für Organismen, deren Stoffwechsel sich deutlich von dem von *E. coli* unterscheidet, könnte die Einbeziehung eines dritten, weit verbreiteten Redoxcofaktors noch größere Vorteile bringen. So nutzen kohlenstofffixierende anaerobe Acetogene (Vertreter der chemolithoautotrophen Mikroorganismen) Ferredoxin (mit einem negativeren Redoxpotential als NAD(P)(H)) als dritten wichtigen Elektronenträger neben NAD(H) und NADP(H). Dieser dritte Redoxcofaktor gibt ihnen zusätzliche Flexibilität, um ihren komplizierten Redox-Stoffwechsel zu balancieren [112].

Einige Aspekte dieser Arbeit sind vergleichbar mit der Studie von Goldford *et al.* [104], in der die Autoren versuchten, Erklärungen für die Redoxcofaktor-Redundanz unter besonderer Berücksichtigung von NAD(H) und NADP(H) zu finden. Für ihre Analysen wurde ebenfalls ein stöchiometrisches Modell von *E. coli* mit integrierten thermodynamischen Nebenbedingungen verwendet. Der grundlegende Unterschied zwischen dem TCOSA-Framework und ihrem Ansatz besteht jedoch darin, dass sie lediglich die thermodynamische Möglichkeit von Stoffwechselflussverteilungen innerhalb dieses Netzwerks analysierten (teilweise mit festen NAD(P)H/NAD(P)⁺-Verhältnissen), während TCOSA die netzwerkweiten thermodynamischen Triebkräfte (MDF) maximierte und einen breiten Bereich von Metabolitkonzentrationen einschließlich Cofaktoren zuließ. Dennoch gibt es einige sich überschneidende Ergebnisse, die wiederum zueinander weitgehend konsistent sind. Beispielsweise verwendeten Goldford *et al.* nicht nur ein stöchiometrisches Modell, sondern auch ein (hypothetisches) minimales Enzymkostenmodell mit randomisierten kinetischen Parametern. Mit letzterem konnte gezeigt werden, dass redundante Redoxcofaktor-Pools höhere thermodynamische Triebkräfte ermöglichen und damit zu geringeren Enzymkosten führen. Dieses Ergebnis stimmt mit den Schlussfolgerungen aus TCOSA überein, die ausschließlich auf Netzwerkstöchiometrie und grundlegenden thermodynamischen Beschränkungen beruhen. Die Schlussfolgerungen von TCOSA deuten in Übereinstimmung damit und darüber hinaus darauf hin, dass das Vorhandensein von zwei Redoxpaaren die maximal erreichbaren thermodynamischen Kräfte innerhalb des Netzwerks maximieren könnte (was wiederum zu niedrigeren Enzymkosten führt). Der wichtigste Unterschied und Fortschritt von TCOSA ist jedoch die Fähigkeit, optimale netzwerkweite NAD(P)(H)-Spezifitäten zu bestimmen, die das thermodynamische Potenzial eines bestimmten metabolischen Reaktionsnetzwerks maximieren. Diese Spezifitäten können dann mit den Wildtyp-Spezifitäten in *E. coli* verglichen werden, was Aufschluss darüber gibt, wie thermodynamische Beschränkungen die Redoxcofaktor-Spezifitäten im metabolischen Netzwerk von *E. coli* beeinflusst haben. Im Vergleich zu den flussbasierten Berechnungen von Goldford *et al.* zeigen die Ergebnisse mit TCOSA, dass eine beträchtliche Anzahl einzelner Cofaktor-Austausche, selbst wenn sie stöchiometrisch nicht wachstumshemmend sind, dennoch einen Einfluss auf die erreichbare thermodynamische Triebkraft innerhalb des Stoffwechselnetzwerks von *E. coli* haben können.

Darüber hinaus gibt es eine beträchtliche Anzahl solcher Austausche, die sich als ungünstig erweisen würden. Diese Beobachtung deutet erneut darauf hin, dass die Entwicklung der Redoxcofaktor-Spezifität in den Stoffwechselreaktionen von *E. coli* durch die Notwendigkeit angetrieben wurde, hohe thermodynamische Potentiale zu ermöglichen.

Als Teil von TCOSA wurde eine Reihe an Werkzeugen und Algorithmen entwickelt. Diese sind allesamt über ein öffentliches GitHub-Repository verfügbar und ermöglichen die Reproduktion der in dieser Arbeit vorgestellten Ergebnisse oder die Durchführung ähnlicher Analysen für andere Stoffwechselnetzwerke. Darüber hinaus ist es möglich, dass sich die mit relevanten Entwicklungen dieser Arbeit für verschiedene andere Anwendungen als wertvoll erweisen könnten. Eine solche Anwendung, die eng mit dem Metabolic Engineering von Zellfabriken für biotechnologische Produktionsprozesse verbunden ist, beinhaltet die Suche nach optimalen NAD(P)(H)-Spezifitäten, die die thermodynamische Triebkraft für ein Zielprodukt maximieren. Andere theoretische Studien haben sich bereits mit der Berechnung optimaler Cofaktor-Austausche zur Steigerung der Produktausbeute befasst [105, 113–117]. TCOSA kann diese Techniken erweitern, indem es thermodynamische Betrachtungen mit einbezieht und so sicherstellt, dass gezielte Änderungen der Cofaktor-Spezifitäten nicht nur die Produktausbeute erhöhen, sondern auch die thermodynamischen Triebkräfte entlang des Weges zum gewünschten Produkt erhöhen. Eine gezielte Änderung der Spezifität von Redoxcofaktoren ist durch verschiedene experimentelle Methoden möglich geworden, darunter direkte Änderungen der Proteinstrukturen [118] sowie adaptiv evolutionäre Techniken (z.B. [119, 120]). Zudem kann sich TCOSA nützlich für die Erklärung weiterer Redoxcofaktor-Redundanzen erweisen, wie etwa der von ATP und GTP [104]. Darüber hinaus könnte die in dieser Arbeit eingeführte Definition des SubMDF potentiell als wertvolles Instrument für künftige thermodynamisch fokussierte Untersuchungen von Stoffwechselnetzwerken oder Teilnetzwerken dienen. Dies ist darauf zurückzuführen, dass die SubMDF die netzwerkweite thermodynamische Möglichkeit einer Lösung beibehält und gleichzeitig eine Fokussierung auf spezifische Reaktionen ermöglicht, die für eine Studie von Interesse sind.

TCOSA motiviert zudem die im nächsten Kapitel entwickelte Methode ASTHERISC.

5 Design mikrobieller Gemeinschaften zur Maximierung der thermodynamischen Triebkraft von Produktsynthesen

5.1 Motivation

In der freien Natur existieren Organismen nur selten als isolierte Populationen einzelner Spezies. Stattdessen bilden sie oft Gemeinschaften, die durch verschiedenartige komplexe Interaktionen zwischen den beteiligten Arten gekennzeichnet sind und oftmals die Evolution der beteiligten Spezies erheblich beeinflussen [121]. Insbesondere mikrobielle Gemeinschaften sind in der Natur weit verbreitet und spielen eine grundlegende Rolle u.a. in der Ökologie, der Bioremediation, den geochemischen Kreisläufen und der menschlichen Gesundheit. Mikrobielle Gemeinschaften umfassen in der Regel eine Vielzahl an Spezies. Speziell unter kontrollierten Laborbedingungen können sie jedoch auch aus mehreren coexistierenden Stämmen *einer* Spezies bestehen, die jeweils deutliche physiologische Unterschiede aufweisen können. Ein Ansatz zur Bildung solcher Einzelspezies- bzw. Multi-Stamm-Gemeinschaften besteht darin, verschiedene genetisch veränderte Stämme mit obligaten gegenseitigen Stoffwechselabhängigkeiten zu konstruieren und gemeinsam zu kultivieren [122].

Speziell mikrobielle Gemeinschaften werden in großem Umfang in biotechnologischen Anwendungen eingesetzt. Ein weit verbreitetes Beispiel ist die Vergärung pflanzlicher Biomasse zu Biogas durch natürlich vorkommende Gemeinschaften. In den letzten Jahren hat sich auch die Erzeugung und Anwendung künstlicher mikrobieller Gemeinschaften verbreitet [123]. Während einige dieser Ansätze mehrere Arten verwenden [124, 125], konzentrieren sich andere auf die Erzeugung von Gemeinschaften aus unterschiedlichen Stämmen einer Art. Erfolgreiche Beispiele hierfür sind 1) eine Gemeinschaft aus zwei *E. coli*-Stämmen für die Umwandlung von Xylan zu Ethanol mit hoher Produkt-zu-Substrat-Ausbeute [126], 2) eine weitere Gemeinschaft aus zwei *E. coli*-Stämmen für die Synthese von Flavonoiden [127] und 3) die Produktion von Anthocyanen, die bisher nicht mikrobiell hergestellt werden konnten, mit einer Gemeinschaft aus vier *E. coli*-Stämmen [128].

Zur theoretischen Untermauerung der Vorteile der praktischen Nutzung mikrobieller Gemeinschaften wird in der Literatur häufig das Konzept der Arbeitsteilung (englisch *division of labor*) verwendet [129–133]. Das Konzept der Arbeitsteilung kann anhand des in Abb. 5.1 dargestellten Beispiels veranschaulicht werden, das sich auf den Fall einer genutzten Art mit mehreren Stämmen konzentriert, für die das Konzept der Arbeitsteilung ebenfalls gelten sollte und die, wie bereits geschrieben, bereits erfolgreich biotechnologisch genutzt wurde. Konkret handelt es sich bei dem Beispiel um einen einfachen Stoffwechselweg, bei dem, noch ohne Berücksichtigung einer Arbeitsteilung, zunächst über die Reaktion R_s ein Substrat S aufgenommen wird. S wird dann durch die vom Enzym E_1 katalysierte Reaktion R_1 in das Zwischenprodukt M umgewandelt. M wiederum wird durch das Enzym E_2 , welches die Reaktion R_2 katalysiert, in das Produkt P umgewandelt, das durch die Reaktion R_p in das Medium ausgeschieden wird. Es ist zu dabei beachten, dass sowohl E_1 als auch E_2 aus zellinternen Ressourcen aufgebaut werden müssen. Diese

internen Ressourcen werden aber auch für andere lebensnotwendige Produkte benötigt. Die gleichzeitige Synthese von E_1 und E_2 aus internen Ressourcen stellt somit eine metabolische Belastung (englisch *metabolic burden*) dar, ein Begriff, der häufig im Zusammenhang mit Arbeitsteilung verwendet wird [131, 134]. Das gängige Konzept der Arbeitsteilung würde hier also wie folgt und wie dargestellt ablaufen (Abb. 5.1): Anstatt den Stoffwechselweg in einer Zelle ablaufen zu lassen, wird eine Gemeinschaft aus zwei genetisch unterschiedlichen Stämmen gebildet. Der erste Stamm führt nur die Reaktionen R_S und R_1 durch und scheidet dann M über die Reaktion $R_{M,ex}$ in das Medium aus. Der zweite Stamm hingegen nimmt M über die Reaktion $R_{M,up}$ auf und führt sowohl R_2 als auch R_P aus. Somit kann der erste Stamm auf E_2 verzichten, und der zweite auf E_1 . Die Zellen der jeweiligen Stämme sollten daher in der Lage sein, einen größeren Anteil ihrer internen Ressourcen für das verbleibende Enzym, E_1 im ersten Stamm und E_2 im zweiten Stamm, zu verwenden. Die daraus resultierende Erhöhung der Enzymkonzentration sollte letztlich den Fluss des Stoffwechselweges erhöhen. Das Konzept der Arbeitsteilung würde also zu einem Flussvorteil durch Verringerung der metabolischen Belastung führen.

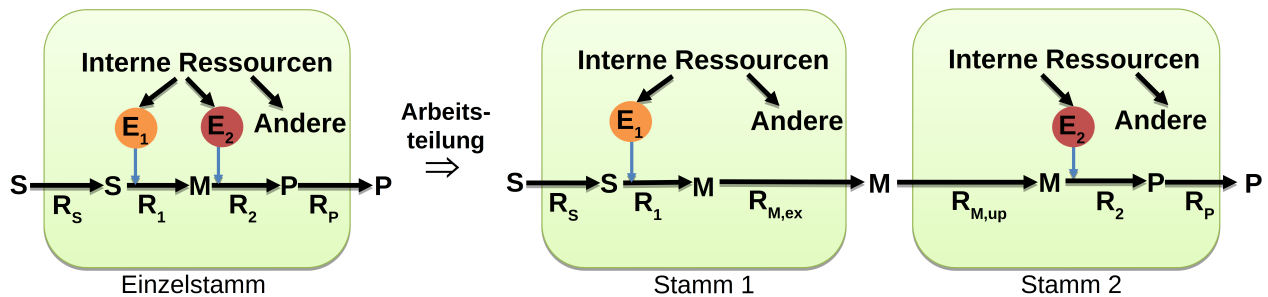


Abbildung 5.1: Visualisierung der in der Literatur verbreiteten Idee der Arbeitsteilung in mikrobiellen Gemeinschaften.

Dieses Konzept der Arbeitsteilung zur Verringerung der metabolischen Belastung funktioniert jedoch nicht, wenn man die Biomasseanteile der einzelnen Stämme betrachtet. Um dies für einen allgemeineren Fall zu verdeutlichen, kann man von einem Stoffwechselweg ausgehen, bei dem, ähnlich wie in Abb. 5.1, ein Substrat S in q Reaktionen zu einem Produkt P umgesetzt wird. Jede Reaktion wird von einem einzigen Enzym katalysiert. Die Reaktionsrate unter dem metabolischen Fließgleichgewicht einer Reaktion i wird als Element r_i des Fließgleichgewichts-Reaktionsratenvektors \mathbf{r} (mit Raten der Einheit $\text{mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$) angegeben. Die Konzentration eines Metaboliten k wird entsprechend als Element c_k des Konzentrationsvektors \mathbf{c} angegeben. \mathbf{r} wird ferner so skaliert, dass die Rate der letzten Reaktion, welche P produziert, 1 ist. Der aus der Skalierung resultierende Vektor heißt $\tilde{\mathbf{r}}$, somit ist $\tilde{r}_p = 1$. Jede Reaktion hat zudem eine, mit z.B. der reversiblen Michaelis-Menten-Kinetik (Gleichung (2.4)) kompatible, assoziierte Reaktionskinetik der Form

$$r_i = [E_i] \cdot k_{cat,i} \cdot f_i(\mathbf{c}) \quad (5.1)$$

mit $[E_i]$ als Konzentration des Enzyms der Reaktion (in $\text{mmol} \cdot \text{gDW}^{-1}$), $k_{cat,i}$ als Wechselzahl des Enzyms der Reaktion und $f_i(\mathbf{c})$ als eine Funktion die alle weiteren kinetischen sowie

thermodynamischen Einflüsse auf die Reaktion enthält, die wiederum von den Metabolitkonzentrationen abhängig sein können (Gleichungen (2.3) bis (2.4)). Die damit resultierenden Enzymkosten für den gesamten betrachteten Stoffwechselweg mit $\tilde{r}_p=1$ sind somit

$$\sum_i^q [E_i] = \sum_i^q \tilde{r}_i / (k_{cat} \cdot f_i(\mathbf{c})). \quad (5.2)$$

Unter der weiteren Annahme einer Beschränkung der maximalen gesamten Enzymmenge Ω gilt ferner

$$\sum_i^q [E_i] \leq \Omega. \quad (5.3)$$

Somit gilt, eingedenk des geforderten Fließgleichgewichts, für die maximale Reaktionsrate der letzten Reaktion, r_p , folgendes für das Verhältnis von Ω zu den Enzymkosten:

$$r_p = \frac{\Omega}{\sum_i^q \tilde{r}_i / (k_{cat} \cdot f_i(\mathbf{c}))} \quad (5.4)$$

Mit einer Gesamtmenge an Biomasse B (in $\text{g} \cdot \text{gDW}^{-1}$) ist ferner der maximale Produktionsfluss J_p (in $\text{mmol} \cdot \text{h}^{-1}$) definiert durch:

$$J_p = B \cdot r_p = B \cdot \frac{\Omega}{\sum_i^q \tilde{r}_i / (k_{cat} \cdot f_i(\mathbf{c}))} \quad (5.5)$$

Der gegebene Stoffwechselweg kann auch in einer mikrobiellen Gemeinschaft stattfinden, beispielsweise in einer mit 2 Stämmen (Stamm 1 und Stamm 2). Diese können sich den Stoffwechselweg untereinander aufteilen mit q_1 Reaktionen für Stamm 1 und q_2 Reaktionen für Stamm 2, wobei $q_1 + q_2 = q$. Der Vektor der Gleichgewichts-Reaktionsraten der gesamten Gemeinschaft \mathbf{r}^c lässt sich somit aus den entsprechenden Vektoren der beiden Stämme als

$$\mathbf{r}^c = \mathbf{r}^1 + \mathbf{r}^2, \quad (5.6)$$

wobei die Reaktionsraten von Stamm 2 in \mathbf{r}^1 , sowie von Stamm 1 in \mathbf{r}^2 , gleich 0 sind. Wichtig ist hierbei, dass in der Gemeinschaft zusätzliche Transportreaktionen für den Austausch von Zwischenprodukten erforderlich sind, zumindest eine Reaktion in jedem Stamm für jeden ausgetauschten Zwischenmetaboliten. Im Folgenden werden diese Transportreaktionen nicht berücksichtigt, d.h. für diese werden keine Enzymkosten angenommen und ihre Reaktionsraten gehen nicht in \mathbf{r}^c ein. In der Realität jedoch würden diese Transportreaktionen typischerweise die Gesamtzymkosten erhöhen und somit den maximal möglichen Fluss mit der Gemeinschaft verringern.

Mit den normierten Vektoren $\tilde{\mathbf{r}}^1$ sowie $\tilde{\mathbf{r}}^2$ erhält man nun

$$\tilde{\mathbf{r}}^c = \tilde{\mathbf{r}}^1 + \tilde{\mathbf{r}}^2 \quad (5.7)$$

mit einer normierten Produktsynthese-Reaktionsrate von

$$\tilde{r}_p^c = 1. \quad (5.8)$$

Unter der Annahme, dass die Metabolitkonzentrationen und die Gesamt-Proteinmenge im Stamm 1 sowie 2 identisch sind, was dem allgemeinen Konzept der vorteilhaften Arbeitsteilung durch geringere metabolische Belastungen nicht widerspricht, lässt sich die maximale Rate, in der \tilde{r}^1 bzw. \tilde{r}^2 pro Gramm an Biomasse der einzelnen Stämme ablaufen kann, als

$$\alpha^1 = \frac{\Omega}{\sum_i^{q_1} \tilde{r}_i^1 / (k_{cat} \cdot f_i(\mathbf{c}))} \quad (5.9)$$

sowie

$$\alpha^2 = \frac{\Omega}{\sum_i^{q_2} \tilde{r}_i^2 / (k_{cat} \cdot f_i(\mathbf{c}))} \quad (5.10)$$

formulieren. Typischerweise, in Abhängigkeit der Enzymkosten einer jeden Reaktion und der Aufteilung des Stoffwechselweges zwischen den beiden Stämmen, sind α^1 und α^2 nicht identisch. Dennoch müssen \tilde{r}^1 sowie \tilde{r}^2 in gleicher Stöchiometrie addiert werden, damit Gleichung (5.8) unter Anbetracht des Fließgleichgewichts valide bleibt. Eine solche passende Aufteilung kann erzielt werden, indem die Biomassen der beiden Stämme so aufgeteilt werden, dass die Gesamtflüsse in den Stämmen zueinander identisch sind, also

$$B^1 \cdot \alpha^1 = B^2 \cdot \alpha^2 = J_p, \quad (5.11)$$

womit aus den Definitionen auch die weitere Definition

$$B = B^1 + B^2. \quad (5.12)$$

folgt. Aus den letzten beiden Gleichungen folgt wiederum:

$$(B - B^2) \cdot \alpha^1 = B^2 \cdot \alpha^2 \Rightarrow B^2 = B \cdot \alpha^1 / (\alpha^2 + \alpha^1) \quad (5.13).$$

Somit erhält man, analog zu Gleichung (5.5) mit einem Einzelstamm, für den maximalen Produktfluss der Gemeinschaft

$$J_p^C = B^1 \cdot \alpha^1 = B^2 \cdot \alpha^2 = \alpha^1 \cdot \alpha^2 \cdot B / (\alpha^2 + \alpha^1) = B / (1/\alpha^1 + 1/\alpha^2). \quad (5.14)$$

Unter Einsetzen von Gleichung (5.9) sowie (5.10) für α^1 sowie α^2 und unter Anbetracht von Gleichung (5.8) erhält man entsprechend folgendes:

$$J_p^C = B^C \cdot \frac{\Omega}{\sum_i^{q_1} \tilde{r}_i^1 / (k_{cat} \cdot f_i(\mathbf{c})) + \sum_i^{q_2} \tilde{r}_i^2 / (k_{cat} \cdot f_i(\mathbf{c}))} = \frac{\Omega}{\sum_i^q \tilde{r}_i^C / (k_{cat} \cdot f_i(\mathbf{c}))} \quad (5.15)$$

Damit ist der Term für den maximalen Produktfluss der Gemeinschaft unter der Annahme gleicher Metabolitkonzentrationen zwischen den Stämmen identisch mit dem entsprechenden Term eines Einzelstammes (Gleichung (5.5)). Damit ist bewiesen, dass mit dem Konzept der Arbeitsteilung

kein Vorteil in der Produktsynthese möglich ist, wenn, dem Konzept nicht widersprechend, identische Metabolitkonzentrationen in den Gemeinschaftskompartimenten angenommen werden. Bemerkenswert ist hierbei, dass kein Vorteil einer Gemeinschaft erkennbar ist, obwohl die metabolischen Kosten (wie z.B. dass Transportenzyme erzeugt werden müssen und der Transport selbst womöglich Energie benötigt) notwendiger Transportvorgänge vernachlässigt wurden.

Wie somit gezeigt, kann eine dem Konzept der Arbeitsteilung folgende Strategie nicht vorteilhaft (oder sogar ungünstig) sein, solange ein potentieller kinetischer Vorteil aufgrund unterschiedlicher Metabolitkonzentrationen in den verschiedenen Stämmen die zusätzlichen Kosten des Metabolittransports nicht aufwiegt. Dieser Punkt wurde bei der Diskussion von Strategien gemäß dem Konzept der Arbeitsteilung bislang selten berücksichtigt.

5.2 Darstellung der Prinzipien von ASTHERISC anhand eines kinetischen Modells

Wie im letzten Abschnitt analytisch gezeigt wurde, ist das allgemeine Konzept der Arbeitsteilung zur Erklärung eines Vorteils von Multi-Stamm-Gemeinschaften bei Produktsynthesen nicht unter allen Annahmen bezüglich der Metabolitkonzentrationen anwendbar. Ergänzend wird in diesem Abschnitt praktisch gezeigt, dass schon unterschiedliche *kritische* Metabolitkonzentrationen zwischen den Kompartimenten dazu führen können, dass eine Gemeinschaft einen Flussvorteil bei einer Produktsynthese erzielen kann.

Ein kleines kinetisches Demonstrationsmodell (Abb. 5.2) soll dieses Prinzip eines potentiellen Flussvorteils einer mikrobiellen Gemeinschaft veranschaulichen. Dieses Modell enthält einen thermodynamisch plausibel gestalteten zentralen Stoffwechselweg. Hierbei wird zunächst der externe Metabolit S_{ex} in die Zelle als S aufgenommen (Reaktion R_1). S wird dann in der Reaktion R_2 zu A und X umgesetzt. A wiederum wird in der Reaktion R_3 , mit dem Metaboliten M als Cofaktor, zu B umgewandelt. B wird dann in der Reaktion R_4 zu C umgesetzt. C wird dann in der Reaktion R_5 mit X zum Produkt P umgewandelt. P wird schließlich in Reaktion R_6 als Endprodukt P_{ex} in das externe Kompartiment ausgeschieden. Der Metabolit M wird durch die Reaktion R_7 durch Aufnahme des externen Metaboliten M_{ex} in die Zelle eingebracht. Zusätzlich kann M durch Reaktion R_8 wieder aus der Zelle als M_{ex} sekretiert werden. Der Metabolit X kann zudem über die Reaktion R_9 aus der Zelle als externer Metabolit X_{ex} sekretiert, und über die Reaktion R_{10} durch Aufnahme von X_{ex} aufgenommen werden.

Dieser Stoffwechselweg wird in 2 Versionen betrachtet, einmal als Einzelstamm, und einmal als eine Gemeinschaft aus zwei Stämmen. Im Einzelstamm läuft der Weg vollständig in einem Kompartiment ab. Im Stamm 1 der Gemeinschaft laufen hingegen nur die Reaktionen R_1 bis R_3 sowie R_7 bis R_{10} ab. Das Produkt B der Reaktion R_3 wird dann in einer zusätzlichen Transportreaktion $R_{TB,1}$ extrazellulär als B_{ex} sekretiert. Die Reaktion $R_{TB,2}$, welche B_{ex} als B aufnimmt, findet dann im Stamm 2 der Gemeinschaft statt. Dort finden wiederum nur die Reaktionen R_4 bis R_6 sowie R_9 und R_{10} statt, die mit der Ausscheidung des Endproduktes P_{ex} enden. Die beiden genannten Stämme 1 und 2 repräsentieren somit eine mikrobielle Gemeinschaft, in welcher der 1. Stamm den Metaboliten B ausscheidet, der 2. Stamm ihn wieder aufnimmt und dann zum Endprodukt P_{ex} weiterverarbeitet. Der Einzelstamm hingegen repräsentiert den Fall, in dem der

Stoffwechselweg in *einer* Zelle ablaufen muss. Diese beiden Varianten des Stoffwechselweges ermöglichen somit einen direkten Vergleich eines Einzelstammes und einer Gemeinschaftsversion davon, hier insbesondere im Hinblick auf den erreichten Fluss zum Endprodukt P_{ex} .

Die modellierte Reaktionskinetik der Reaktionen R_1 bis R_6 sowie $R_{TB,1}$ und $R_{TB,2}$ entspricht der bereits dargestellten (Gleichung (2.4)) zerlegten reversiblen Michaelis-Menten-Kinetik [36] und ist in Tab. A1 im Anhang vollständig beschrieben. Die Zufuhr- und Ausfuhrreaktionen R_7 bis R_{10} haben eine Flussrate, die durch einen einzelnen konstanten Parameter und der Konzentration ihres jeweiligen assoziierten Metaboliten M sowie X bestimmt wird (siehe hierzu zusätzlich die Differentialgleichungen der Metabolitkonzentrationen des Modells in Tab. A2 im Anhang). Die im Rahmen der im folgenden beschriebenen Analyse als Ausgangspunkt verwendeten kinetischen Startparameter (Tab. A3 im Anhang) jeder Reaktion wurden zufällig gewählt.

Entscheidend für die Darstellung des genannten thermodynamischen Prinzips eines potentiellen Gemeinschaftsvorteils sind sowohl die Festlegung der Substrat- und Produktkonzentrationen als auch die Verteilung der $\Delta_r G'^{\circ}$ -Werte im Modell (Abb. 5.2). Bei den Metabolitkonzentrationen werden $[S_{ex}]$ sowie $[P_{ex}]$ auf einen konstanten Wert von 4 M bzw. 2 M gesetzt. Die Konzentrationen der internen Metabolite M bzw. X werden wesentlich aus den Raten ihrer Zufuhr- und Ausfuhrreaktionen R_7 bis R_{10} beeinflusst. Diese Reaktionen repräsentieren somit ein Einstellen der Konzentrationen der Metabolite M bzw. X durch weitere, in diesem Modell nicht weiter ausgeführte, Stoffwechselwege der Zelle. Alle weiteren Metabolitkonzentrationen erhalten in jeder Runde einen Startwert von 10^{-6} M.

Hinsichtlich der $\Delta_r G'^{\circ}$ ist zu beachten, dass sowohl der $\Delta_r G'^{\circ}$ der Reaktion R_2 , bei der X produziert wird, als auch der $\Delta_r G'^{\circ}$ der Reaktion R_4 , bei der X verbraucht wird, einen positiven $\Delta_r G'^{\circ}$ von +4 kJ/mol aufweisen. Es besteht somit ein Zielkonflikt bezüglich der Konzentration von X zur Optimierung des $\Delta_r G'$ bzw. damit des maximal möglichen Produktflusses (Gleichung (2.7)): Für die Reaktion R_2 sollte nämlich hierfür die Konzentration von X niedrig, für die Reaktion R_4 diese hingegen hoch sein. Eine hier vorgeschlagene Lösung dieses Problems, die letztlich zu einer höheren Produktionsrate führen soll, ist die bereits beschriebene Aufteilung des Stoffwechselweges in mehrere Kompartimente bzw. Stämme. Die Transportreaktionen R_1 , R_6 , $R_{TB,1}$ und $R_{TB,2}$ haben alle einen $\Delta_r G'^{\circ}$ von 0 kJ/mol, sodass für einen ablaufenden Transport ein Konzentrationsgefälle notwendig ist und der Transport nicht thermodynamisch gesondert begünstigt wird, was der Gemeinschaft einen künstlichen Vorteil verschaffen könnte. Darüber hinaus haben die Reaktionen R_3 sowie R_4 einen negativen $\Delta_r G'^{\circ}$ von -5 kJ/mol und sind daher thermodynamisch nicht stark limitierend.

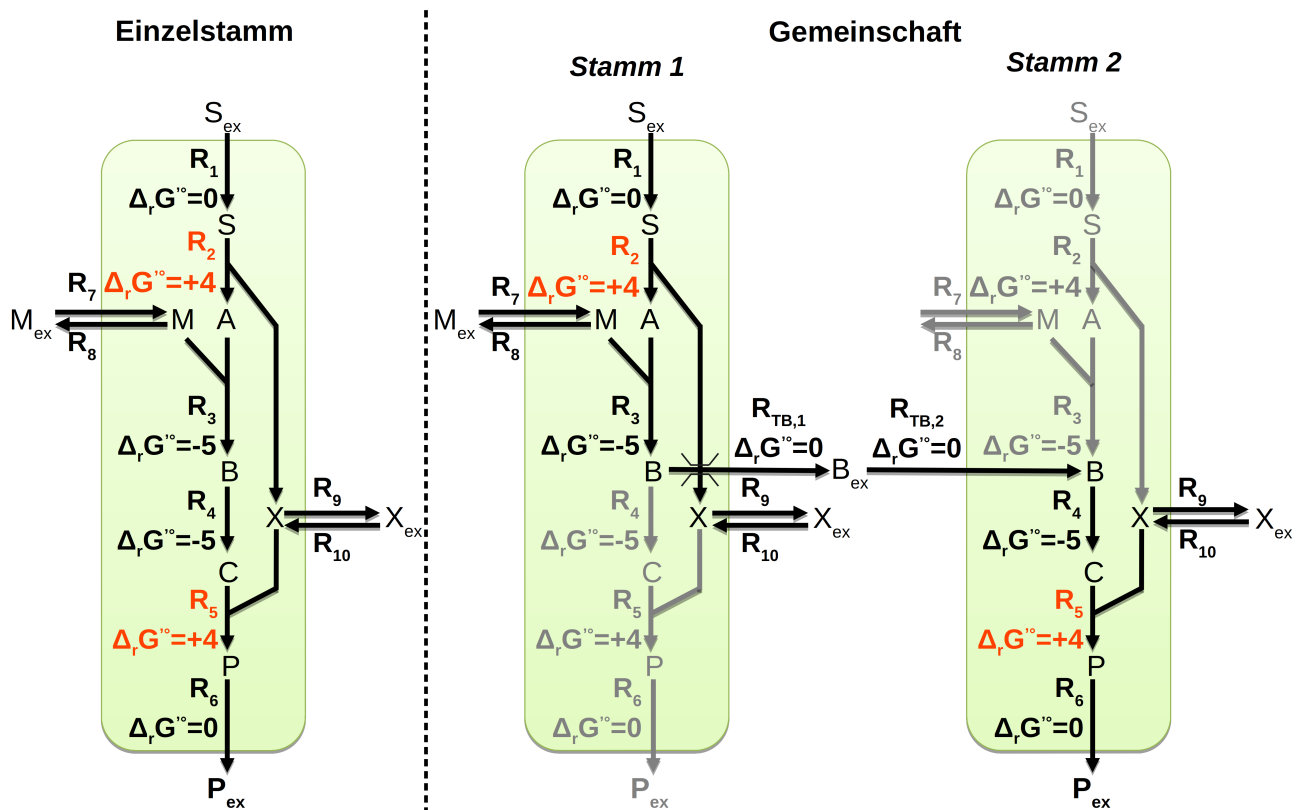


Abbildung 5.2: Visualisierung des beispielhaften Pathways zur Entwicklung eines Prinzips thermodynamischer Gemeinschaftsvorteile. Der Pathway läuft hierbei einmal in einem Einzelstamm, und einmal aufgeteilt in einer Gemeinschaft zweier Stämme ab.

Zur Analyse des kinetischen Modells wurde eine randomisierte Studie durchgeführt, um die P_{ex} -Produktionsraten zwischen dem Einzelstamm und der Gemeinschaft zu vergleichen. Die Studie bestand aus 5.000 Einzelrunden. Vor jeder Runde wurden alle kinetischen Parameter mit Ausnahme der $\Delta_r G^\circ$ und den Enzymkonzentrationen variiert (siehe auch Tab. A3 im Anhang). Dabei wurde zunächst durch einen stochastischen Münzwurf entschieden, ob der jeweilige Parameter erhöht oder verringert werden sollte. Anschließend wurde der Parameter auf einen Zufallswert zwischen dem ursprünglichen Ausgangswert und dem um den Faktor 100 erhöhten bzw. erniedrigten Wert gesetzt. Die Parameter sowohl in der Einzelspezies als auch in der Gemeinschaft wurden hierbei um den gleichen Faktor verändert und blieben somit in jeder Runde jeweils untereinander identisch. Die einzige, und für die nachfolgenden Analysen wichtige, Ausnahme bildeten hierbei die kinetischen Parameter der X_{ex} -assoziierten Reaktionen R_9 sowie R_{10} , welche in jeder Runde nur im Einzelstamm und im Stamm 1 der Gemeinschaft identisch gehalten wurden, aber sich im Stamm 2 der Gemeinschaft hiervon unterscheiden können. Hiermit wird simuliert, dass die Gemeinschaft in der Lage ist, in seinem Stamm 2 eine andere Konzentration von X über R_9 sowie R_{10} , und damit den wie beschrieben nicht weiter im Modell betrachteten weiteren Stoffwechselwegen, einstellen zu können. Bei passender Änderung der kinetischen Parameter von R_9 sowie R_{10} könnte die Gemeinschaft somit in der Lage sein, den im letzten Absatz beschriebenen Zielkonflikt, um das Einstellen einer für die P_{ex} -Produktion vorteilhaften X -Konzentration zu lösen. Mit dem so veränderten Satz kinetischer Parameter in jeder Runde wurde in einer Simulation ein Fließgleichgewichts-Zustand ermittelt. War diese Zustandsberechnung nicht erfolgreich, oder der resultierende Fluss von R_6 im Einzelstamm

und/oder der Gemeinschaft kleiner als 0,1, wurde die Runde verworfen. Andernfalls wurde sie akzeptiert und in die unten aufgeführten Statistiken aufgenommen.

Folgende Resultate ergaben sich nach Ausführung der 5.000 Runden (Tab. 5.1): Während 2.910 dieser Runden aufgrund eines nicht auffindbaren Fließgleichgewichts oder eines zu geringen Flusses verworfen wurden, verblieben 2.090 akzeptierte Runden. In 1196 bzw. ca. 57,2 % der akzeptierten Runden konnte ein P_{ex} -Produktionsratenvorteil der Gemeinschaft gegenüber dem Einzelstamm festgestellt werden, während in 894 bzw. ca. 42,8 % die Gemeinschaft keinen Vorteil aufwies. Die Mechanik hinter den Gemeinschaftsvorteilen lässt sich auf das Auflösen des schon beschriebenen Zielkonfliktes bei der Konzentration von X zurückführen (Abb. 5.3): Nur wenn die resultierende Konzentration von X im Stamm 2 der Gemeinschaft ($[X_2]$) höher ist als die entsprechende Konzentration im Stamm 1 ($[X_1]$) kann die Gemeinschaft auch einen Vorteil erzielen. Ist $[X_2]$ hingegen kleiner als $[X_1]$, konnte in keinem Falle ein Gemeinschaftsvorteil ermittelt werden. Somit muss $[X_2]$ als Konzentration eines Substrats einer thermodynamisch beschränkenden Reaktion tatsächlich hochreguliert werden. Unter Betrachtung der, in diesem kinetischen Modell verwendeten, zerlegten reversiblen Michaelis-Menten-Kinetik (Gleichung (2.4)) werden somit bei einem Gemeinschaftsvorteil der Sättigungsterm (κ) und/oder der thermodynamische Term (γ) der Gemeinschaftsreaktionen durch das Einstellen optimierter Metabolitkonzentrationen im Vergleich zum Einzelstamm erhöht. Damit wird gleichzeitig deren beschränkender Einfluss auf den maximalen Fluss (V^+) gesenkt.

Tabelle 5.1: Zentrale Resultate der vorliegenden randomisierten Studie. Das Vorliegen des Flussvorteils einer Gemeinschaft ist definiert als höhere Produktionsrate des Endproduktes P_{ex} im Gemeinschaftsmodell im Vergleich zum Einzelstamm des kinetischen Beispielmodells bei einer gegebenen Parametrisierung. $[X_1]$ steht für die Konzentration von X im 1., $[X_2]$ für die im 2. Stamm der Gemeinschaft. In keiner der akzeptierten Parametrisierungen war $[X_1]$ identisch zu $[X_2]$.

	(Akzeptierte) Parametrisierungen	Mit Flussvorteil der Gemeinschaft	Ohne Flussvorteil der Gemeinschaft
<u>Insgesamt</u>	2.090	1.196	894
Mit $[X_2] > [X_1]$	1.306	1.196	110
Mit $[X_2] < [X_1]$	784	0	784

Interessanterweise konnte jedoch in 110 Runden kein Vorteil der Gemeinschaft festgestellt werden, obwohl $[X_2]$ größer als $[X_1]$ eingestellt war. Dies ist auf den Nachteil der Gemeinschaft zurückzuführen, dass die Konzentrationsentwicklung von $[X_1]$ bzw. $[X_2]$ von den Flüssen jeweils dreier Reaktionen (R_2 , R_9 und R_{10} im Stamm 1 bzw. R_9 , R_{10} und R_5 im Stamm 2) abhängt, während im Einzelstamm die Konzentrationsentwicklung von X von 4 Reaktionsflüssen (R_2 , R_5 , R_9 und R_{10}) abhängt. Dies ermöglicht einerseits prinzipiell den beschriebenen Mechanismus des Flussvorteils der Gemeinschaft. Bei einer ungünstig hohen Konzentrationseinstellung von $[X_1]$ (über die Raten von R_2 , R_9 und R_{10}) ist es jedoch andererseits möglich, dass der Einzelstamm sein $[X]$ über die zusätzliche X-assoziierte Reaktion (R_5) optimierter einstellen und einen Flussvorteil gegenüber der Gemeinschaft erzielen kann. Mechanistisch führt in diesem Falle die ungünstig hohe Einstellung

von X_1 in der Gemeinschaft dazu, dass die Konzentration von S ($[S]$) im Vergleich zum Einzelstamm erhöht werden muss. Dies wiederum senkt die maximale Flussrate von R_1 im 1. Gemeinschaftsstamm und damit die Aufnahme des Substrats S_{ex} , was letztlich die Produktionsrate in der Gemeinschaft senkt. Tatsächlich ist in allen 110 Fällen, in denen trotz eines günstigen $[X_2]:[X_1]$ -Verhältnisses kein Gemeinschaftsvorteil besteht, $[S]$ in der Gemeinschaft immer höher als im Einzelstamm. Trotz dieser potentiell ungünstigen Einstellungsmöglichkeit hatte die Gemeinschaft einen Flussvorteil in insgesamt 1196 von 1306 und damit ca. 91,6 % aller Fälle, in denen in der Gemeinschaft $[X_2]$ größer als $[X_1]$ eingestellt wurde. In all diesen 1196 Fällen wurde $[S]$ auch immer in der Gemeinschaft niedriger als im Einzelstamm eingestellt.

Unter Berücksichtigung dieser Ergebnisse konnten somit folgende zwei Schlussfolgerungen gezogen werden: 1) Ergänzend zu der analytischen Betrachtung im letzten Abschnitt, dass eine stammabhängige Variation der Metabolitkonzentrationen für einen Flussvorteil einer Gemeinschaft notwendig ist, wurde im kinetischen Modell gezeigt, dass hierfür konkret die Konzentrationen kritischer Metabolite variabel sein müssen. Im vorliegenden Falle betraf dies den Metaboliten X, dessen Konzentration in der randomisierten Studie zwischen den Stämmen der Gemeinschaft unterschiedlich eingestellt werden musste (Tab. 5.1; Abb. 5.3). Nur wenn hierbei die Konzentration von X im Stamm 2 der Gemeinschaft größer war als die von X im Stamm 1, konnte die Gemeinschaft einen Vorteil erzielen. Somit wird (ähnlich wie im vorherigen Abschnitt analytisch nahegelegt) gezeigt, dass ein allgemeines Konzept zur Beschreibung eines Gemeinschafts-Flussvorteils auch zumindest das Einstellen unterschiedlicher kritischer Metabolitkonzentrationen für jeden Stamm einer Gemeinschaft erlauben muss. 2) Im Gegensatz zur analytischen Betrachtung im letzten Abschnitt werden nun explizit auch Transportreaktionen mit dem Erfordernis eines Konzentrationsgefälles einbezogen, womit gezeigt werden konnte, dass auch mit solchen Reaktionen eine Gemeinschaft tatsächlich einen Flussvorteil erzielen kann.

Alle hier dargestellten Berechnungen wurden mit Python 3.8 sowie Tellurium 2.1 [81] durchgeführt. Einzelne Tests des kinetischen Beispielmmodells wurden zusätzlich in COPASI 4.33 [80] durchgeführt. Die Ergebnisse und Skripte sind unter folgendem Repository verfügbar:

<https://github.com/Paulocracy/kinetic-modelling> oder <https://doi.org/10.5281/zenodo.10556533>

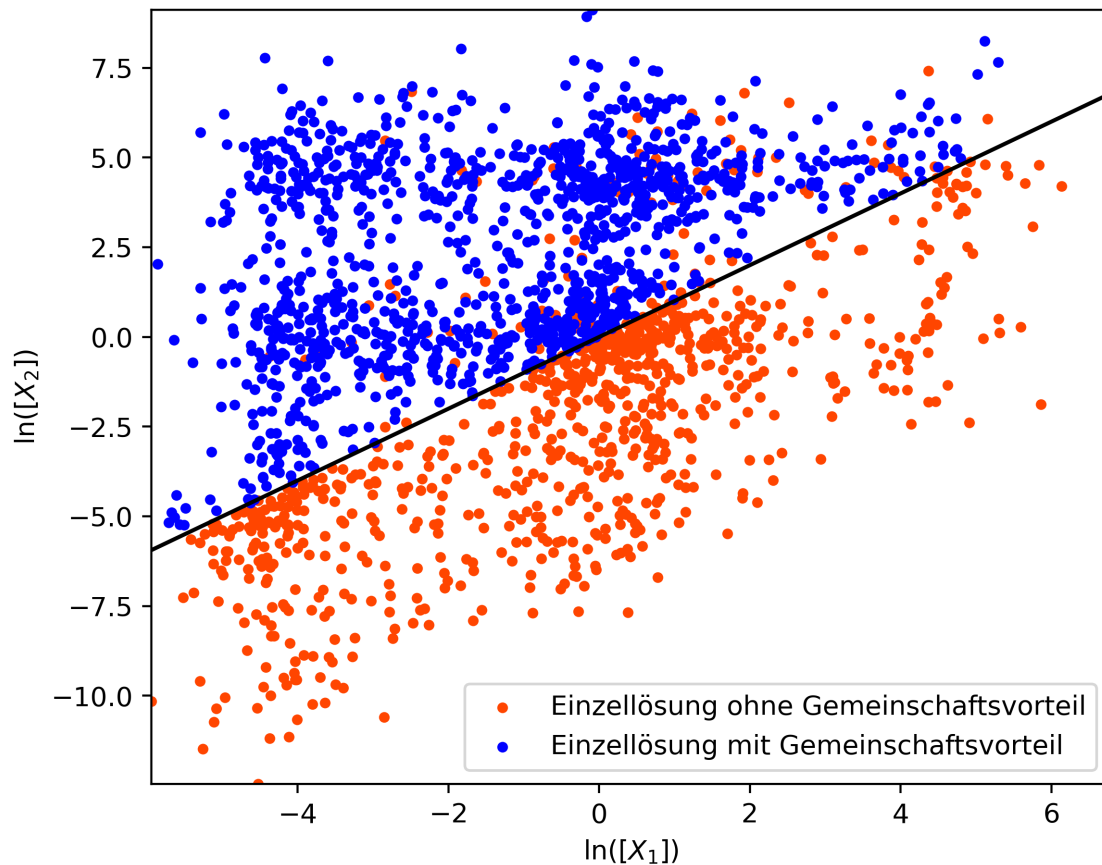


Abbildung 5.3: Vergleich der logarithmierten Konzentrationen von X im 1. und 2. Gemeinschaftsstamm in jeder Einzellösung der randomisierten Studie, aufgeschlüsselt nach dem Auftreten eines Gemeinschaftsvorteils bei der Produktionsrate von P_{ex} . Auf der schwarzen Linie ist $\ln([X_1]) = \ln([X_2])$. Blaue Punkte können in dieser Abbildung rote Punkte überdecken.

5.3 Die ASTHERISC-Methode

Theoretische Grundlage

Das im letzten Abschnitt anhand eines kinetischen Modells vorgestellte Konzept eines Vorteils mikrobieller Gemeinschaften, bestehend aus mehreren Stämmen einer Spezies, für einen Produktsynthesefluss soll nun mithilfe des bereits eingeführten Konzepts der maximal-minimal netzwerkweiten thermodynamischen Triebkraft (Abb. 2.1D), kurz MDF, auf constraint-basierte stöchiometrische Modelle übertragen und zum Design solcher Gemeinschaften verwendet werden.

Wie im letzten Abschnitt veranschaulicht, besteht in einer Gemeinschaft ein nutzbarer Freiheitsgrad, um eine Gesamtproduktionsrate im Vergleich zu einem Einzelstamm zu verbessern, darin, unterschiedliche Metabolitkonzentrationen in den einzelnen Stämmen der Gemeinschaft einstellen zu können. Im schon beschriebenen kinetischen Beispielmodell (Abb. 5.2) gelingt dies, unter den dort getroffenen Annahmen, in allen Fällen eine Verbesserung der Produktionsrate von P_{ex} durch ein stammabhängiges Einstellen des Metaboliten X , bei dem die Konzentration von X im Stamm 2 der Gemeinschaft im Vergleich zum Stamm 1 erhöht sein muss. Auch wenn in einzelnen Fällen dies nicht zu einem Produktionsvorteil führte, trat ein solcher Vorteil bei allen Fällen mit

einem solchen Konzentrationsverhältnis in ca. 91,6 % dieser Fälle auf (Tab. 5.1). Somit ist ein solches Konzentrationsverhältnis augenscheinlich eine plausible Heuristik zur Vorhersage eines möglichen Produktionsvorteils einer Gemeinschaft. Der Zusammenhang mit dem MDF (siehe Kapitel 2.1.6) ergibt sich dadurch, dass die mit X assoziierten, und für die Produktion von P_{ex} notwendigen, Reaktionen R_2 sowie R_5 durch ihre hohen $\Delta_r G^\circ$ -Werte thermodynamische Flaschenhälse darstellen. Da zudem bei R_2 der Metabolit X ein Produkt und bei R_5 ein Substrat ist, lässt sich im Falle des Einzelstammes die Konzentration von X nicht erhöhen bzw. erniedrigen, ohne dass die thermodynamische Triebkraft von R_2 bzw. von R_5 und somit der MDF erniedrigt wird. Bei einer Aufspaltung des Syntheseweges von S zu P in zwei Stämme, wie in Abb. 5.2 visualisiert, kann hingegen die Konzentration von X je nach Stamm MDF-optimierend eingestellt werden. Im Stamm 2 kann entsprechend die Konzentration vom R_2 -Produkt X gesenkt, und im Stamm 3 die vom R_5 -Substrat X erhöht werden, ohne dass die thermodynamische Triebkraft einer anderen Reaktion erniedrigt wird. Somit kann die Produktion von S zu P_{ex} über zwei Stämme einen höheren MDF als mit nur einem Stamm aufweisen.

Der im letzten Absatz postulierte MDF-Vorteil der Gemeinschaft lässt sich durch eine Anwendung von OptMDFpathway (Abschnitt 2.1.6) auf den Einzelstamm und die Gemeinschaft aufzeigen: Hierzu wird aus dem kinetischen Modell ein stöchiometrisches constraint-basiertes Modell abgeleitet. Alle Reaktionen, die entsprechend Abb. 5.2 aktiv sein sollen, haben freie Flussgrenzen im Bereich $(-\infty; \infty)$ und weisen, wo vorhanden, zueinander jeweils identische $\Delta_r G^\circ$ -Werte auf. Genauso wie im kinetischen Modell (Tab. A1 im Anhang) hat der Metabolit P eine Konzentration von 2 M und S eine von 4 M, während alle anderen Metabolite einen breiten Konzentrationsbereich von 10^{-6} M bis 100 M aufweisen können. Als Gaskonstante R und Temperatur T werden zudem die genannten Standardwerte verwendet (siehe Abschnitt 2.1.6). Um nun den MDF bei der Synthese von P aus S aus einem Einzelstamm zu berechnen, wird ein positiver Fluss in Reaktion R_6 des 1. Stammes gefordert. Zur Berechnung des MDF der Gemeinschaft wird ein entsprechender Fluss in R_6 des 2. Stammes gefordert. Die Anwendung von OptMDFpathway ergibt nun mit diesen Einstellungen einen MDF des Einzelstammes von ca. 2,52 kJ/mol. Die Gemeinschaft hingegen erreicht einen deutlich höheren MDF von ca. 4,42 kJ/mol und hat damit einen MDF-Vorteil von ca. 1,9 kJ/mol. Um zu zeigen, dass – wie im kinetischen Modell – das stammabhängige Einstellen der Konzentration von X eine entscheidende Rolle beim Erreichen des hohen MDF der Gemeinschaft hat, bietet sich eine Analyse von Metabolit-Konzentrationsverhältnissen (ähnlich zu Gleichungen (4.6) bis (4.10)) an. Hierbei wird das Verhältnis von $[X_1]:[X_2]$, also zwischen dem Metaboliten X im Stamm 1 sowie 2 der Gemeinschaft, unter Einhaltung des Gemeinschafts-MDF maximiert. Als Resultat ergibt sich, dass das maximale $[X_1]:[X_2]$ -Verhältnis ca. 0,00028 betragen muss. Wenn, als weitere Berechnung, nur der niedrigere MDF des Einzelstammes gefordert wird, beträgt das maximale $[X_1]:[X_2]$ -Verhältnis ca. 0,13 und ist somit ebenfalls deutlich unter 1. Ähnlich wie in allen Fällen mit Gemeinschaftsvorteil im kinetischen Modell senkt die Gemeinschaft zur Einhaltung höherer MDF somit tatsächlich die Konzentration des R_2 -Produktes X_1 im Vergleich zur erhöhten Konzentration des R_4 -Substrates X_2 . Dieses Beispiel legt nahe, dass das Potential einer Gemeinschaft zum Erzielen eines möglichen thermodynamischen Produktsynthesevorteils durch das Auffinden von MDF-Vorteilen einer Gemeinschaft identifiziert werden kann. Die hier

beschriebene OptMDFpathway-Berechnung ist im genannten GitHub-Repository des kinetischen Beispielmotells auffindbar (siehe vorherigen Abschnitt).

Aufbauend auf letzterem Erkenntnis wurde im Rahmen dieser Arbeit der Algorithmus ASTHERISC (Algorithmic Search of Thermodynamic Advantages in Single-Species Communities) entwickelt. Dieser kann alle Produkte in einem Stoffwechselmodell identifizieren, bei denen eine Gemeinschaftsversion eines metabolischen Modells einer einzelnen Spezies zu einem MDF-Vorteil der Gemeinschaft bei der Produktsynthese führt. Diese Arbeit konzentriert sich dabei auf die Produktsynthese an sich, ohne Betrachtung von Wachstum, und demonstriert die Anwendbarkeit der Methode durch die Identifizierung von Multi-Stamm-Gemeinschaften aus *E. coli*, die den MDF der Produktsynthese für eine breite Palette von Produkten maximieren können.

Aufbau der Gemeinschaftsmotelle

Die mit ASTHERISC zu verwendenden Stoffwechselmodelle müssen speziell konfiguriert werden (ähnlich Abb. 2.2). Zunächst müssen, wie in der in dieser Arbeit verwendeten Formulierung von OptMDFpathway erforderlich, alle reversiblen Reaktionen in irreversible Reaktionen zerlegt werden. Außerdem muss, ausgehend von einem Einzelspeziesmodell, für jedes mit ASTHERISC zu testende Produkt eine Austauschreaktion im externen Kompartiment eingefügt werden. Dieses Einzelspeziesmodell wird dann durch Vervielfachung (z.B. Verdopplung) des Einzelspezieskompartiments in ein Gemeinschaftsmotell überführt. Dabei werden auch alle austauschbaren Produkte als Metaboliten in das Austauschkompartiment der Gemeinschaft aufgenommen. Gleiches gilt für alle essentiellen Substrate und ausgewählte Standardnebenprodukte. Für diese werden zusätzlich Austauschreaktionen mit der Umwelt in das Austauschkompartiment des Gemeinschaftsmotells integriert. Im Gegensatz zu typischen Gemeinschaftsmotellen geht ASTHERISC von einer Gemeinschaftswachstumsrate von 0 h^{-1} aus. Damit konzentriert sich ASTHERISC ausschließlich auf die Synthesewege der untersuchten Produkte und deren thermodynamisches Potential. Dementsprechend gelten auch die ursprünglichen Flussgrenzen (Gleichung (2.10)) aus dem Einzelstamm weiter, und die Anteile der einzelnen Stämme werden ignoriert, da der Biomasseanteil eines Stammes keinen Einfluss auf das (nicht betrachtete) Wachstum und damit die Flussgrenzen eines anderen Stammes haben kann.

Einbau thermodynamischer Nebenbedingungen

Jede Analyse mit ASTHERISC basiert zum einen auf den Nebenbedingungen der CBM mit Gemeinschaftsmotellen (Gleichungen (2.34) bis (2.40)), wobei für jeden Stamm der Gemeinschaft ein identischer Biomasseanteil und eine Wachstumsrate von 0 h^{-1} angenommen wird, und zum anderen auf OptMDFpathway (Gleichungen (2.25) bis (2.31)), wobei statt der Parameter einer Einzelspezies die Parameter der Gemeinschaft verwendet werden. Ein Unterschied zur bisher betrachteten OptMDFpathway-Methode besteht in der Verallgemeinerung der Nebenbedingungen für Metabolitkonzentrationsverhältnisse (Gleichung (2.27)). Statt eines festen Verhältnisses kann in der ASTHERISC-Variante ein minimales und ein maximales Verhältnis über

$$x_i - x_j \geq \ln(h_{min}) \tag{5.16}$$

sowie

$$x_i - x_j \leq \ln(h_{max}) \quad (5.17)$$

gesetzt werden. Ferner werden im Rahmen des gesamten ASTHERISC-Algorithmus (siehe nächsten Unterabschnitt) folgende Zielfunktionen verwendet:

OptMDF1: MDF-Optimierung

Hier wird, wie in der ursprünglichen OptMDFpathway-Formulierung und unter den angegebenen Nebenbedingungen, die MDF-Variable B maximiert, also konkret

$$\begin{aligned} & \text{maximiere } B \\ & \quad r, x, z, B \\ & \text{u. d. N. (2.25) bis (2.30), (2.34) bis (2.40), (5.16) bis (5.17)} \end{aligned}$$

OptMDF2: MDF-Optimierung mit Einschränkung der Austauschreaktionen

Diese Variante basiert auf OptMDF1. Es wird jedoch eine zusätzliche Nebenbedingung hinzugefügt, um die Gesamtzahl der aktiven „Extra“-Austauschreaktionen zu begrenzen. Als „Extra“-Austauschreaktionen werden alle Austauschreaktionen mit dem Austauschkompartiment definiert, die Metabolite austauschen, die weder essentielle Substrate für das Wachstum liefern noch typische Sekretionsprodukte darstellen. Konkret lautet die Nebenbedingung

$$\sum_{i \in EX} z_i \leq c \quad (5.18)$$

mit EX als die Menge aller Indizes der Extra-Austauschreaktionen und c als die Obergrenze der aktiven Extra-Austauschreaktionen. Somit lautet eine Optimierung von B mit OptMDF1

$$\begin{aligned} & \text{maximiere } B \\ & \quad r, x, z, B \\ & \text{u. d. N. (2.25) bis (2.30), (2.34) bis (2.40), (5.16) bis (5.18)} \end{aligned}$$

OptMDF3: Lösung mit geforderter unterer MDF-Grenze

Dies entspricht OptMDF1, jedoch wird anstatt einer Optimierung von B lediglich ein minimaler MDF-Wert γ per

$$B \geq \gamma \quad (5.19)$$

verlangt. Somit folgt eine „Optimierung“ ohne Zielfunktion, lediglich eine die Nebenbedingungen erfüllende Lösung wird gesucht.

OptMDF4: Minimierung der Flusssumme

Diese Variante basiert auf OptMDF3, wobei eine Zielfunktion für das Minimieren der Gesamtsumme der Reaktionsflüsse per

$$\begin{aligned} & \text{minimiere } \sum_{i=1}^m r_i \\ & \quad r, x, z, B \\ & \text{u. d. N. (2.25) bis (2.30), (2.34) bis (2.40), (5.16) bis (5.19)} \end{aligned}$$

festgelegt wird. Da alle Reaktionen in ASTHERISC-kompatiblen Modellen irreversibel sind (siehe vorherigen Unterabschnitt), entspricht dies einer Minimierung der absoluten Reaktionsflüsse. Diese Zielfunktion ähnelt damit der Parsimonious FBA, kurz pFBA, bei der nach Optimierung der FBA-Zielfunktion ebenfalls die Summe der Flüsse minimiert wird [135].

OptMDF5: Konzentrations-Variabilitätsanalyse

Diese basiert auf OptMDF3 und führt eine Konzentrations-Variabilitätsanalyse durch, wobei ein minimaler MDF gefordert werden kann (Gleichung (5.20)). Hierfür wird für jeden Metaboliten j dessen Konzentration zunächst über

$$\underset{r, x, z, B}{\text{minimiere } x_j}$$

u. d. N. (2.25) bis (2.30), (2.34) bis (2.40), (5.17) bis (5.18), (5.20)

minimiert und wie folgt maximiert:

$$\underset{r, x, z, B}{\text{maximiere } x_j}$$

u. d. N. (2.25) bis (2.30), (2.34) bis (2.40), (5.17) bis (5.18), (5.20)

Der ASTHERISC-Algorithmus

Der ASTHERISC-Algorithmus verwendet die im letzten Unterabschnitt beschriebenen Methoden, um in einem ASTHERISC-kompatiblen Gemeinschaftsmodell, das wie beschrieben aus einer Einzelspezies erstellt wurde, für alle ausgewählten Produkte zu bestimmen, ob es einen MDF-Gemeinschaftsvorteil gibt. Wenn dies der Fall ist, werden zusätzlich relevante Informationen zum Verständnis des Vorteils gesammelt. Der Algorithmus läuft wie folgt ab (Algorithmus 5.1): Als Vorbereitung wird zunächst aus einer gegebenen Spezies wie beschrieben das Gemeinschaftsmodell mit zwei oder mehreren Stämmen erzeugt. Programmtechnisch wird dann das Einzelstammmodell durch die Deaktivierung aller Austauschreaktionen in allen Kompartimenten außer dem ersten rekonstruiert. Für jeden Zielmetaboliten werden nun folgende Schritte durchgeführt: 1) Die Austauschreaktion des Zielmetaboliten aus dem Austauschkompartiment in die Umgebung wird in allen Modellen aktiviert. 2) Anschließend wird mit dem Einzelstammmodell die maximale Produkt-Substrat-Ausbeute berechnet (siehe auch Gleichung (2.34) und die zugehörige Optimierung). 3) Mit der vorgegebenen minimalen Ausbeute wird der MDF im Einzelstamm- und im Gemeinschaftsmodell berechnet. Für das Gemeinschaftsmodell kann eine Obergrenze für den zusätzlichen Austausch festgelegt werden. Nur wenn der MDF im Gemeinschaftsmodell höher ist als im Einzelstammmodell, werden die folgenden Schritte durchgeführt: 4) Zunächst wird die maximale Ausbeute des Gemeinschaftsmodells beim MDF berechnet. 5) Damit wird dann der Flussvektor des Gemeinschaftsmodells mit der minimalen Flusssumme beim MDF und der dabei auftretenden maximalen Ausbeute berechnet. 6) Optional kann anschließend eine Konzentrations-Variabilitäts-Analyse der Metaboliten mit den in diesem Flussvektor aktiven Reaktionen und 7) eine Berechnung der thermodynamischen Flaschenhalse durchgeführt werden.

```

# Setze Parameter #
# Das Modell mit nur einem Stamm
set einzelstammModell
# Anzahl der Stämme im Gemeinschaftsmodell
set anzahlStämme
# Minimal notwendiger MDF-Vorteil des Gemeinschaftsmodells
set minimalerMdfVorteil
# Minimal notwendiger Anteil der maximalen Ausbeute des
# Einzelspezies-Modells mit dem Gemeinschaftsmodell
set minimalerAusbeuteAnteil
# Maximale Anzahl Austausch im Gemeinschaftsmodell (kann  $\infty$  sein)
set maximaleAustausche
# Liste der Ziel-Metabolite für die ASTHERISC durchgeführt wird
set zielMetabolite

# Erstellen des Gemeinschaftsmodells mit der gegebenen Stammanzahl.
gemeinschaftsModell = erstelleGemeinschaftsModell(
    einzelstammModell, anzahlStämme
)
for zielMetabolit in zielMetabolite:
    # Schritt 1: Erlaube (temporär) einen Export des
    # Ziel-Metaboliten in die Umgebung
    einzelstammModellZielMetabolit =
        erhalteEinzelstammModellMitZielMetaboliten(
            einzelstammModell, zielMetabolit
        )
    gemeinschaftsModellZielMetabolit =
        erhalteGemeinschaftsModellMitZielMetaboliten(
            gemeinschaftsModell, zielMetabolit
        )

    # Schritt 2: Berechne die maximale Produkt-zu-Substrat-
    # Ausbeute im Einzelstamm-Modell
    maximaleAusbeuteEinzelstamm =
        berechneMaximaleAusbeute(
            einzelstammModellZielMetabolit
        )
    minimalNotwendigeAusbeute = minimalerAusbeuteAnteil *
        maximaleAusbeuteEinzelstamm

    # Schritt 3: Berechne den MDF sowohl im Einzelstamm-
    # als auch im Gemeinschaftsmodell
    mdfEinzelstamm = OptMDF1(
        einzelstammModellZielMetabolit,
        minimalNotwendigeAusbeute
    )
    mdfGemeinschaft = OptMDF2(
        einzelstammModellZielMetabolit,
        minimalNotwendigeAusbeute,
        maximaleAustausche
    )
    # Wenn der MDF des Gemeinschaftsmodells keinen
    # genügenden MDF-Vorteil im Vergleich zum Einzelstamm-
    # Modell bietet, fahre mit dem nächsten Produkt fort
    minimumMdf = mdfEinzelstamm + minimalerMdfVorteil
    if mdfGemeinschaft < minimumMdf:
        continue

```

Algorithmus 5.1: Pseudocode des ASTHERISC-Algorithmus.


```

# Schritt 4: Berechne die maximale Produkt-zu-Substrat-
# Ausbeute des Gemeinschaftsmodells bei dessen MDF bei Synthese
gemeinschaftsMaxAusbeute@MDF =
  MaxAusbeute@MaxMDFmitOptMDF3 (
    gemeinschaftsModellZielMetabolit,
    maximaleAustausche,
    mdfGemeinschaft
  )

# Schritt 5: Berechne einen Flussvektor mit der minimalen
# absoluten Flusssumme bei der maximalen Produkt-zu-
# Substrat-Ausbeute sowie dem MDF bei Produktsynthese
gemeinschaftsMinFlussVektor@MaxAusbeute@MDF =
  OptMDF4 (
    gemeinschaftsModellZielMetabolit,
    gemeinschaftsMaxAusbeute@MDF,
    maximaleAustausche,
    optMdfGemeinschaft
  )
aktiveReaktionen@MinFluss@MaxAusbeute@MDF =
  aktiveReaktionen (
    gemeinschaftsMinFlussVektor@MaxAusbeute@MDF
  )
# Schritt 6 (optional): Berechne eine Konzentrations-
# Variabilitätsanalyse mit den aktiven Reaktionen des
# Flussvektors aus Schritt 5
metabolitKonzentrationsBereiche =
  erhalteMetabolitKonzentrationsBereicheMitOptMDF5 (
    gemeinschaftsModellZielMetabolit,
    aktiveReaktionen@MinFluss@MaxAusbeute@MDF
  )

# Schritt 7 (optional): Berechne die thermodynamischen
# Flaschenhälse mit den aktiven Reaktionen des Flussvektors
# aus Schritt 5
flaschenhalse = berechneFlaschenhälse (
  einzelstammModellZielMetabolit,
  aktiveReaktionen@MinFluss@MaxAusbeute@MDF
)

```

(Fortsetzung von Algorithmus 5.1)

5.4 Implementierung von ASTHERISC

Programmatisch ist der ASTHERISC-Algorithmus in Form zweier miteinander kompatibler Programmpakete, `CommModelPy` sowie das `ASTHERISC Package` implementiert (Abb. 5.4).

CommModelPy

Mit `CommModelPy` kann ein Gemeinschaftsmodell aus einem oder mehreren verschiedenen Einzelspeziesmodellen erstellt werden. Dabei kann eine Linearisierung entweder durch feste Speziesanteile oder durch eine feste Wachstumsrate gemäß den Gleichungen (2.38) sowie (2.39) erfolgen. Darüber hinaus kann `CommModelPy` direkt ASTHERISC-kompatible Modelle erzeugen, etwa mit Modifikationen wie der Nichtbeachtung von Wachstum und Verwendung von Identifizierern, die für das `ASTHERISC Package` (siehe unten) geeignet sind. Dies unterscheidet `CommModelPy` von ähnlichen Funktionalitäten der Pakete `CarveMe` [33] und `micom` [136].

CommModelPy ist in Python implementiert und basiert auf COBRApy [10]. Metabolische Modelle können damit unter anderem im SBML-Format [78] geladen und gespeichert werden. CommModelPy kann unter folgendem GitHub-Repository eingesehen werden, wo auch alle in diesem Kapitel verwendeten Gemeinschaftsmodelle und deren Generierungsskripte zu finden sind:

<https://github.com/klamt-lab/CommModelPy>

ASTHERISC Package

Das ASTHERISC Package ist in MATLAB geschrieben, verwendet Funktionen von *CellNetAnalyzer* [11] und führt die Schritte 1 bis 7 des ASTHERISC-Algorithmus (Algorithmus 5.1) aus. Als Eingaben benötigt das Paket ein mit CommModelPy erstelltes ASTHERISC-kompatibles Gemeinschaftsmodell sowie eine Liste in Form einer JSON-Datei.

Der Programmcode sowie die im Rahmen dieses Kapitels erstellten Daten und zugehörigen Skripte sind im folgenden GitHub-Repository abgelegt:

<https://github.com/klamt-lab/astheriscPackage>

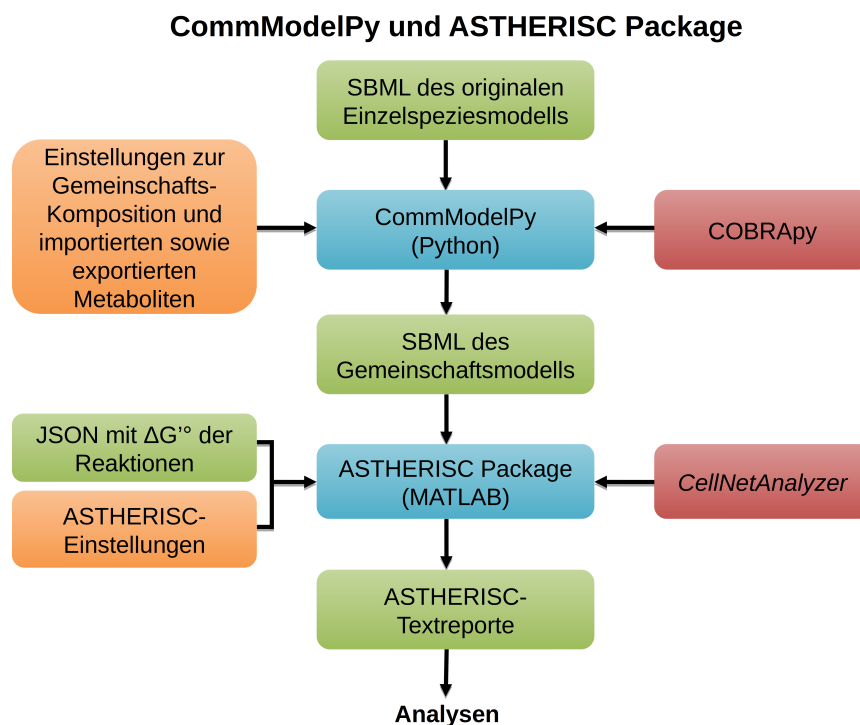


Abbildung 5.4: Übersicht der kombinierten Nutzung von CommModelPy und dem ASTHERISC Package. Orange Kästen stehen für Benutzereinstellungen, grüne für generierte oder gegebene Daten, rote für zentrale Abhängigkeiten der Programmpakete und blaue für die Programme selbst.

5.5 Anwendung von ASTHERISC an Gemeinschafts-Modellen von *E. coli*-Stämmen

Als beispielhafte realistische Anwendung von ASTHERISC wurde systematisch untersucht, welche spezifisch designte Gemeinschaften von *E. coli*-Stämmen die Synthese von Metaboliten thermodynamisch optimieren können.

Datenquellen

Das originale *iML1515*-Modell wurde aus der BiGG-Datenbank [77] als SBML-Datei bezogen. Das aus *iJO1366* [70] abgeleitete reduzierte Modell *EColiCore2* (ECC2) wurde direkt aus dessen Veröffentlichung bezogen [71]. Alle hier genutzten $\Delta_r G^\circ$ wurden mithilfe eines Python-Skriptes ermittelt, welches die *equilibrator*-API [45] nutzt.

Genutzte Softwareversionen

Die hier gezeigten Berechnungen nutzten Python 3.7, MATLAB 2018a, IBM CPLEX 12.9, *CellNetAnalyzer* 2019.2 sowie COBRApy 0.18.

Ermittlung der Standard-Gibbs-Energie-Werte

Die Ermittlung der $\Delta_r G^\circ$ -Werte in den Basismodellen ECC2 sowie *iML1515* wurde wie folgt durchgeführt: Zunächst wurde, soweit möglich, für jede Reaktion in ECC2 sowie *iML1515* ein Wert mithilfe der Python-API des *eQuilibrator* [45] berechnet. Da die API zum Zeitpunkt der Arbeiten an ASTHERISC noch nicht mit anderen Identifizierern kompatibel war, wurden die Reaktionen und deren Metabolite der API über bereits in *iML1515* vorhandene Identifizierer von *MetaNetX* [76] eingegeben. Als Einstellungen für die Berechnung der wurden eine Temperatur von 298,15 K, eine Ionenstärke von 0,25 M, ein pH-Wert von 7,0 und ein μMg von 3,0 verwendet. Die von der API berechneten Unsicherheiten wurden nicht berücksichtigt, sodass eine Ungenauigkeit von 0 kJ/mol angenommen wurde, um mögliche thermodynamisch unplausible Zyklen zu vermeiden. Alle Reaktionen mit einer $\Delta_r G^\circ$ -Unsicherheit größer als 10.000 kJ/mol, oder für welche die API keinen $\Delta_r G^\circ$ berechnen konnte, wurden im Gemeinschaftsmodell deaktiviert, d.h. die untere und obere Flussgrenze wurden auf 0 $\text{mmol}\cdot\text{gDW}^{-1}\cdot\text{h}^{-1}$ gesetzt. Einzige Ausnahme waren alle Reaktionen, deren BiGG-Identifizierern auf „tpp“, „t1pp“ sowie „tex“ enden, da diese Endungen für zentrale Transportreaktionen stehen. Diese Transportreaktionen wurden ohne thermodynamische Limitierung und ohne Änderung der Flussraten im Gemeinschaftsmodell belassen. Zusätzlich wurde, da zum Zeitpunkt der Arbeiten an ASTHERISC die entsprechende Funktionalität im *eQuilibrator*-Paket nicht vorhanden war, separat vom *eQuilibrator* ein $\Delta_r G^\circ$ -Korrekturfaktor für die zentralen membrangebundenen Reaktionen der ATP-Synthase (BiGG-Identifizierer ATPS4), der NADH-Dehydrogenasen (NADH16, NADH17 und NADH18) sowie der Cytochromoxidasen (YTBDpp, CYTBD2pp und CYTBO3_4p) angewendet. Dieser Faktor basiert auf den Formeln in (136), die wiederum auf (137) basieren, und berücksichtigt den Einfluss des mit den Reaktionen stattfindenden Protonentransports. Konkret wird der Faktor für eine Reaktion mittels folgender Formel berechnet:

$$C_i = c_i \cdot F \cdot \Delta\Psi - 2,3 \cdot h_i \cdot R \cdot T \cdot \Delta pH \quad (5.20)$$

Hierbei steht c_i für die im Ablauf der Reaktion stattfindende Ladungsdifferenz, $\Delta\Psi$ für das elektrochemische Potential der Membran (hier, wie in [137], -130 mV), F für die Faraday-Konstante, h_i für die Anzahl der transportierten Protonen, R für die universelle Gaskonstante, T für die Temperatur und ΔpH für die pH-Differenz zwischen den betroffenen Kompartimenten. Als Werte für c_i sowie h_i wurden die Daten aus dem Anhang von [137] verwendet.

Durch dieses Verfahren wurden ca. 18 % der Reaktionen in ECC2 und ca. 24 % der Reaktionen in *iML1515* aufgrund des Fehlens eines berechneten $\Delta_r G^\circ$ deaktiviert.

Erstellung der Gemeinschaftsmodelle

Zunächst wurden mit CommModelPy die Basismodelle ECC2 sowie *iML1515* mit geeigneten Austauschreaktionen präpariert. Hierzu wurde für jeden Metaboliten eine Austauschreaktion in die Umgebung eingeführt, entweder mit periplasmatischen Variante des Metaboliten oder, falls diese Variante nicht existiert, mit der cytosolischen Variante. Bei der Kombination der Basismodelle zu den (weiter unten beschriebenen) Gemeinschaftsmodellen dienen diese Exportreaktionen als Austauschreaktionen des Stammkompartiments mit dem Austauschkompartiment.

Für die unten beschriebene Kombination der Basismodelle zu den im Folgenden beschriebenen Gemeinschaftsmodellen wurden folgende Einstellungen vorgenommen: Als $\Delta_r G^\circ$ -Werte wurden für alle Reaktionen in den Stammkompartimenten die im vorherigen Unterabschnitt gewählten Werte verwendet. Zusätzlich wurde für alle Austauschreaktionen von den Stammkompartimenten zu den Austauschkompartimenten ein Standard- $\Delta_r G^\circ$ von 0 kJ/mol angenommen. Dies erfordert ein Konzentrationsgefälle für einen thermodynamisch plausiblen passiven Transport. Als Substrat wurde stets Glucose mit einer festen Aufnahmerate von $1 \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ verwendet. Zusätzlich wurden einige Metabolite, darunter essentielle und Standardfermentationsprodukte, als „Standardmetabolite“ definiert (Tab. A4 im Anhang). Für diese wird zum einen die Nutzung von Austauschreaktionen in Gleichung (5.19) nicht berücksichtigt. Zum anderen können sie unabhängig vom gerade untersuchten Zielmetaboliten im ASTHERISC-Algorithmus (Algorithmus 5.1) aufgenommen und/oder sekretiert werden.

Mit all den beschriebenen Einstellungen wurden sodann mit CommModelPy folgende Gemeinschaftsmodelle erstellt:

1. *ecolicore2double*: Dies ist eine zweifache Kombination des wie beschrieben präparierten ECC2-Modells. Dies resultierte in insgesamt, im gesamten Gemeinschaftsmodell betrachtet, 2211 Reaktionen sowie 1398 Metabolite.
2. *ecolicore2triple*: Entspricht *ecolicore2double* mit der einzigen Änderung, dass es sich um eine dreifache Kombination des präparierten ECC2-Modells handelt. Es enthält insgesamt 3105 Reaktionen und 1885 Metaboliten.
3. *iML1515double*: Eine zweifache Kombination des wie beschrieben präparierten *iML1515*. Es verfügt über 8265 Reaktionen und 4924 Metabolite.

Aufgrund der hohen Anzahl an Reaktionen und Metabolite und damit hohen Rechenintensität des ASTHERISC-Algorithmus mit iML1515double wurde auf das Erstellen eines iML1515triple-Modells verzichtet.

Weitere thermodynamische Einstellungen

Je nach Vorkommen in den Kompartimenten und der Art der Metabolite wurden unterschiedliche Konzentrationsbereiche festgelegt (siehe Gleichung (2.26)). Standardmäßig lag dieser Bereich sowohl für cytoplasmatische als auch für periplasmatische Metabolite zwischen 1 μM und 20 mM. Dies entspricht dem Bereich aus der Originalpublikation von OptMDFpathway [15]. Für CO_2 wurde, aufgrund seiner geringen intrazellulären Löslichkeit, ein Konzentrationsbereich von 1 μM bis 1 mM festgelegt. Da die Löslichkeit von Molekülen im exoplasmatischen Raum nicht so streng begrenzt ist, wurde der Standardkonzentrationsbereich für alle Metabolite dort auf 1 μM bis 10 M festgelegt. Darüber hinaus wurde speziell für Wasser (BiGG-Identifizierer h2o) und Protonen (h) in allen Kompartimenten eine feste Konzentration von 1 M festgelegt, da die Berechnungen der $\Delta_r G''$ mit der eQuilibrator-API bereits eine „aktive“ Wasserkonzentration voraussetzen und die Protonenkonzentration bereits über die pH-Einstellung festgelegt und enthalten ist.

Zusätzlich wurden mehrere physiologisch relevante Bereiche an Konzentrationsverhältnissen festgelegt. Dies gilt für [ATP]:[ADP] mit einem Bereich von 3 bis 10, [ADP]:[AMP] von 0,5 bis 2, [NAD⁺]:[NADH] von 3 bis 10 und [NADPH]:[NADP⁺] von 3 bis 10. Diese Werte orientieren sich an den festen Bereichen der Originalpublikation von OptMDFpathway [15], wurden jedoch gelockert, um einen größeren Lösungsraum für diese Bereiche zu erhalten.

Einige ASTHERISC-spezifische Einstellungen wurden ebenfalls vorgenommen. Der minimale MDF-Vorteil einer Gemeinschaft gegenüber einem einzelnen Stamm wurde auf 0,2 kJ/mol festgelegt. Wenn der Vorteil geringer war, wurde die entsprechende Lösung nicht als Gemeinschaftsvorteil registriert, wodurch minimale Vorteile ausgeschlossen wurden. Der minimale MDF, der von einer Gemeinschaft erreicht werden musste, war 0 kJ/mol. Dies schließt alle thermodynamisch unmöglichen Lösungen aus. Es ist jedoch immer noch möglich, dass ein Einzelstamm in seiner Lösung einen negativen MDF-Wert erreicht, während er in der Gemeinschaft aufgrund eines Gemeinschaftsvorteils positiv wird. Um numerische Probleme zu vermeiden, wurde als Mindestwert für den Reaktionsfluss $10^{-6} \text{ mmol} \cdot \text{gDW}^{-1} \cdot \text{h}^{-1}$ verwendet, um die entsprechende Reaktion als aktiv zu bezeichnen. Aus dem gleichen Grund wurde ein intrazellulärer Metabolit nur dann als „produzierbar“ angesehen, wenn die minimale Produkt-zu-Substrat-Ausbeute 10^{-6} mol/mol betrug. Die Genauigkeit der Näherungsroutinen für die MDF und die maximal erreichbare Produkt-Substrat-Ausbeute wurde auf 0,1 kJ/mol bzw. 0,0 mmol/mmol festgelegt. Außerdem wurde eine maximale Rechenzeit von 1000 s für jede einzelne Berechnung festgelegt. Alle Berechnungen, die länger dauerten, wurden verworfen und ein Warnhinweis wurde in die endgültigen Textberichte aufgenommen.

Mit den Gemeinschaftsmodellen getestete Szenarien

Mit dem wie oben beschrieben (aufgrund fehlender $\Delta_r G''$) verringerten Satz an aktiven Reaktionen und zusätzlich, um potentielle numerische Probleme zu vermeiden, einer geforderten minimalen Produktausbeute von 10^{-6} , gab es 161 produzierbare Metabolite in *ecolicore2double* sowie *ecolicore2triple*, und 254 in *iML1515double*. Für jeden der produzierbaren Metaboliten wurden in jedem dieser drei Community-Modelle 8 verschiedene Szenarien betrachtet, die sich aus vier verschiedenen Ausbeuten (40 %, 60 %, 80 % und 98 % der maximalen Ausbeute des jeweiligen Metaboliten) und zwei verschiedenen Grenzen für die maximale Anzahl (zusätzlicher) Austauschreaktionen (∞ oder maximal 9 zusätzliche Austauschreaktionen) ergeben. Insgesamt werden somit 24 Szenarien betrachtet.

Um Lösungen mit nur marginalen MDF-Verbesserungen auszuschließen, wurden zudem ein minimaler MDF-Vorteil der Gemeinschaft von 0,2 kJ/mol verlangt. Zusätzlich wurde ein positiver OptMDF für die Gemeinschaftslösung (nicht für eine assoziierte Einzelstammlösung) verlangt, um Ergebnisse zu erhalten, die thermodynamisch relevant sind.

Resultate

Die Anwendung des ASTHERISC-Algorithmus (Algorithmus 5.1) ergab mit den beschriebenen Gemeinschaftsmodellen sowie Szenarien folgende Resultate:

Unter der bereits beschriebenen Mindestausbeute von Produkt zu Substrat ergaben sich zunächst sowohl für *ecolicore2double* als auch für *ecolicore2triple* insgesamt 161 produzierbare Metabolite und für *iML1515double* 254 Metabolite.

Für alle 24 getestete Szenarien, die zuvor erläutert wurden, und somit für alle Modelle, die Anzahl der zusätzlichen Austauschreaktionen sowie die minimalen Produktausbeuten, konnte ASTHERISC eine vielfältige Menge an produzierbaren Metaboliten mit einem Vorteil der Gemeinschaft für den MDF im Vergleich zum Einzelstamm identifizieren. Der Anteil der produzierbaren Metabolite mit einem solchen Vorteil lag zwischen Werte von 15,75 % und 40,55 % (Tab. 5.2). Erwartungsgemäß führte die Beschränkung der Anzahl der zusätzlichen Austauschreaktionen auf 9 fast immer zu einer Verringerung dieses Anteils. In den Fällen, in denen dies nicht eintrat, führte lediglich die eingestellte maximale Rechenzeit zu einer höheren Anzahl nicht identifizierter Metaboliten mit MDF-Vorteil, da MILP-Berechnungen mit unendlich vielen erlauben Extra-Austauschreaktionen aufgrund des größeren Lösungsraumes rechentechnisch aufwendiger sind als mit einer begrenzten Anzahl Extra-Austauschreaktionen.

Tabelle 5.2: Prozentsatz (und Anzahl) aller produzierbarer Metabolite bei denen ASTHERISC einen MDF-Vorteil der Gemeinschaft im Vergleich zum Einzelstamm identifizieren konnte. Hierbei werden, wie in den genannten Einstellungen gegeben, nur Vorteile größer als 0,2 kJ/mol registriert.

Modell (Anzahl erlaubter Extra-Austausche)	Geforderte minimale Produktausbeute (als % der maximalen Ausbeute)			
	40 %	60 %	80 %	98 %
<i>ecolicore2double</i> (9)	22,36 % (36)	21,74 % (35)	21,12 % (34)	21,12 % (34)
<i>ecolicore2double</i> (∞)	23,6 % (38)	35,4 % (57)	27,95 % (45)	31,06 % (50)
<i>ecolicore2triple</i> (9)	21,12 % (34)	19,88 % (32)	17,39 % (28)	19,25 % (31)
<i>ecolicore2triple</i> (∞)	19,25 % (31)	31,06 % (50)	27,33 % (44)	29,81 % (48)
<i>iML1515double</i> (9)	17,72 % (45)	15,75 % (40)	21,26 % (54)	25,59 % (65)
<i>iML1515double</i> (∞)	18,5 % (47)	20,08 % (51)	25,59 % (65)	40,55 % (103)

In den Fällen, in denen die Gemeinschaft einen MDF-Vorteil erzielte, lagen die Vorteile je nach Szenario zwischen 0,2 kJ/mol und 6,28 kJ/mol (Tab. 5.3). Die durchschnittlichen Vorteile in diesen Fällen betragen 0,63 kJ/mol bis 2,0 kJ/mol für *ecolicore2double* und *ecolicore2triple* und 0,91 kJ/mol bis 1,94 kJ/mol für *iML1515double*. Dabei traten die höchsten absoluten MDF-Vorteile durchgängig bei der höchsten geforderten maximalen Ausbeute von 98 % auf. Dies kann dadurch erklärt werden, dass bei hohen Ausbeuten weniger produktsynthesefähige Stoffwechselwege zur Verfügung stehen, während bei niedrigeren Ausbeuten mehr solcher Wege und damit mehr potentiell MDF-vorteilhafte Alternativen in einem Stamm vorhanden sind. Darüber hinaus wurden mehrere Metabolite identifiziert, deren Produktsynthese im Einzelstamm thermodynamisch unmöglich ist (negativer MDF), die aber in der Gemeinschaft wieder möglich wird (positiver MDF). Bei einer minimalen Produkt-Substrat-Ausbeute von 98 % waren dies 7 Metabolite mit *ecolicore2double* und *ecoli2triple* und 3 Metabolite mit *iML1515double*.

Tabelle 5.3: Statistik der ermittelten MDF-Vorteile der Gemeinschaft im Vergleich zum Einzelstamm in kJ/mol für alle 24 berechneten Szenarien. Die Zahlen geben den minimalen/mittleren/maximalen MDF-Vorteil an.

Modell (Anzahl erlaubter Extra-Austausche)	MDF-Vorteil der Gemeinschaft in kJ/mol			
	40 %	60 %	80 %	98 %
<i>ecolicore2double</i> (9)	0,32/0,94/2,76	0,2/0,64/3,13	0,29/0,67/3,24	0,27/2,0/5,49
<i>ecolicore2double</i> (∞)	0,32/1,29/3,42	0,26/1,02/5,07	0,22/1,04/5,07	0,23/1,77/6,28
<i>ecolicore2triple</i> (9)	0,32/0,83/2,76	0,22/0,63/3,03	0,29/0,7/3,24	0,5/1,85/5,49
<i>ecolicore2triple</i> (∞)	0,29/1,42/3,42	0,26/1,09/4,71	0,22/0,93/4,71	0,23/1,81/6,28
<i>iML1515double</i> (9)	0,26/0,92/2,95	0,33/1,0/2,79	0,23/0,93/2,66	0,22/1,58/5,55
<i>iML1515double</i> (∞)	0,26/0,97/3,04	0,26/0,91/2,79	0,3/0,91/2,64	0,22/1,94/6,04

Die Anzahl der jeweils benötigten Extra-Austauschreaktionen betrug je nach Szenario zwischen 1 und 51 (Tab. 5.4). Dabei ist zu beachten, dass in den Fällen, in denen die maximale Anzahl dieser Reaktionen auf 9 beschränkt wurde, dies zu einem Austausch von maximal 4 Metaboliten führte, unter Nutzung von jeweils maximal 4 Reaktionen vom ersten Stamm in das Austauschkompartiment, 4 Reaktionen vom Austauschkompartiment in den zweiten Stamm und zusätzlich eine Reaktion, die den Zielmetaboliten vom zweiten Stamm in das Austauschkompartiment sekretiert. Letztere Reaktion ist für die Synthese zwingend notwendig und erklärt auch den Minimalwert von 1, bei welchem nur Standardmetabolite (Tab. A4 im Anhang) ausgetauscht werden, die hier nicht in die Statistik eingehen.

Einige Metabolite kamen in den Lösungen besonders häufig als Austauschmetabolite vor. Beispielsweise traten Glycerol-3-phosphat (BiGG-Identifizierer glyc3p), Fructose-6-phosphat (f6p) sowie Oxalacetat (oaa) besonders häufig in *ecolicore2double* sowie *ecolicore2triple* auf, während Metabolite wie Dihydroxyaceton (dha), L-Aspartat (asp_L) sowie Acetaldehyd (acald) häufig in *iML1515double* auftraten. Die meisten dieser Metabolite sind direkt oder über kurze Wege indirekt an der Glycolyse beteiligt. Dies deutet auf zentrale thermodynamische Flaschenhälse hin, die für die Synthese vieler Zielprodukte relevant zu sein scheinen.

Darüber hinaus traten einige diskrete MDF-Werte in jedem Szenario besonders häufig auf. Häufig betrifft dies Produkte, die im metabolischen Netzwerk nahe beieinander liegen, sodass die MDF-Werte ihrer Synthese durch die gleichen thermodynamischen Flaschenhälse begrenzt wurden. Dies deutet darauf hin, dass es einige zentrale Flaschenhälse gibt, die eine große Anzahl möglicher Produktsynthesen thermodynamisch beschränken.

Tabelle 5.4: Statistiken über die Anzahl der Extra-Austauschreaktionen der Gemeinschaft in den 24 berechneten Szenarien für alle Fälle, in denen die Gemeinschaft einen MDF-Vorteil gegenüber dem einzelnen Stamm aufweist. Die angegebenen Zahlen stellen die minimale/mittlere/maximale Anzahl an Extra-Austauschreaktionen dar.

Modell (Anzahl erlaubter Extra-Austausche)	Anzahl verwendeter Extra-Austauschreaktionen			
	40 %	60 %	80 %	98 %
<i>ecolicore2double</i> (9)	3/5,39/8	4/6,34/7	4/6,41/9	3/6,03/9
<i>ecolicore2double</i> (∞)	3/16,66/29	7/15,88/31	6/13,71/29	8/15,88/41
<i>ecolicore2triple</i> (9)	3/4,47/7	3/5,88/7	4/5,89/7	3/5,29/9
<i>ecolicore2triple</i> (∞)	3/18,13/38	3/14,8/36	5/13,36/34	4/16,83/51
<i>iML1515double</i> (9)	1/5,18/7	3/5,05/9	1/5,11/9	2/5,14/9
<i>iML1515double</i> (∞)	3/15,28/33	3/14,16/27	3/14,11/37	3/14,34/28

Um einen beispielhaften Mechanismus zu identifizieren, der dem Vorteil von Gemeinschafts-MDF zugrunde liegt, wurde zudem eine mit ASTHERISC errechnete exemplarische Lösung der Produktsynthese von 3-Deoxy-D-mannoctulose-8-phosphat (kdo8p) mit *ecolicore2double*,

einer minimalen Produkt-zu-Substrat-Ausbeute von 98 % sowie einer maximalen Anzahl von 9 Austauschreaktionen näher betrachtet (Abb. 5.5). Die minimal zu erreichende Ausbeute beträgt hier 0,732795 mol/mol.

Mit den genannten Einstellungen wurde für die Synthese von kdo8p einen MDF von 2,27 kJ/mol mit dem Einzelstamm und 2,87 kJ/mol mit der Gemeinschaft erzielt. Hierbei erreichte der Einzelstamm eine Ausbeute von 0,7455 mol/mol, die Gemeinschaft hingegen eine Ausbeute von 0,73311 mol/mol (98,04 % des Einzelstammes). Das Gemeinschaftsmodell nutzte hierbei 66 aktive Reaktionen, von denen 22 Austauschreaktionen darstellen, 28 interne Reaktionen des ersten Stammes und 16 des zweiten Stammes waren. 17 der 22 Austauschreaktionen stellten dabei Standard-Austauschreaktionen (Tab. A4 im Anhang) dar. Zwei Extra-Austauschreaktionen dienten dem Austausch von Dihydroxyaceton (dha) vom ersten in den zweiten, und zwei weitere für den Austausch von 6-Phospho-D-Gluconat (6pgc) vom zweiten in den ersten Stamm (Abb. 5.5). Die verbleibende Austauschreaktion war das Ausscheiden von kdo8p durch den zweiten Stamm. In der Lösung des Einzelstammes waren 42 Reaktionen aktiv, wovon 14 Standard-Austauschreaktionen darstellten.

Um zu verstehen, wie die Gemeinschaft ihren MDF-Vorteil erreichte, ist es hilfreich, die thermodynamischen Flaschenhalsreaktionen im Einzelstammmodell zu betrachten. Drei dieser Reaktionen waren durch gemeinsame Metabolite miteinander verbunden und bilden einen zentralen Teil der Glycolyse (rot in Abb. 5.5): 1) die Fructose-Bisphosphatase-Aldolase-Reaktion (BiGG-Identifizier FBA; mit einem $\Delta_r G^\circ$ von +22,4 kJ/mol), die Glyceraldehyd-3-Phosphat (g3p) sowie Dihydroxyacetonphosphat (dhap) produziert, 2) die Triose-Phosphat-Isomerase (TPI; $\Delta_r G^\circ$ von +5,6 kJ/mol), die dhap zu g3p umsetzt und 3) der Glyceraldehyd-3-phosphat-Dehydrogenase-Reaktion (GAPD; $\Delta_r G^\circ$ von +4,6 kJ/mol), die g3p zu 1,3-Disphosphoglycerat (13dpg) umsetzt. Wie bereits in [47] sowie [15] festgestellt, stellten diese drei Reaktionen einen verteilten thermodynamischen Flaschenhals dar aufgrund a) ihres jeweils positiven $\Delta_r G^\circ$ und b) ihres Zielkonfliktes hinsichtlich der Konzentration von g3p bei einem MDF-Optimierungsprozess. Letzterer war darauf zurückzuführen, dass sowohl FBA als auch TPI thermodynamisch von einer niedrigen Konzentration von g3p profitieren, während GAPD von einer höheren Konzentration profitieren würde. Ein weiterer thermodynamischer Flaschenhals war ein Phosphatsympporter (Plt2rpp) mit einem $\Delta_r G^\circ$ -Wert von 0 kJ/mol. Da dieser für die weiteren Schlussfolgerungen irrelevant ist, wird er hier nicht weiter betrachtet.

In der Gemeinschaft wurde der Zielkonflikt um g3p, wie weiter unten erläutert, in dem oben genannten Flaschenhals-Dreieck gelöst. Der Mechanismus war wie folgt (Abb. 5.5): Um kdo8p über die kdo8p-Synthase-Reaktion (KDOPS) synthetisieren zu können, wurden für diesen Schritt sowohl D-Arabinose-5-Phosphat (ara5p) als auch Phosphoenolpyruvat (pep) als Substrate benötigt. In der Einstammlösung wurde ara5p über einen 6pgc nutzenden Stoffwechselweg synthetisiert, der keinen thermodynamischen Flaschenhals darstellt. Der Metabolit pep hingegen wurde durch die Glycolyse gebildet, wofür die bereits erwähnten 3 Flaschenhalsreaktionen FBA, TPI und GAPD benötigt wurden. Dieses Flaschenhals-Dreieck wurde nun in der gemeinsamen Lösung dadurch aufgelöst, dass die g3p-verbrauchende Reaktion GAPD nur im ersten Stamm aktiv ist, während die

g3p-produzierenden Reaktionen FBA sowie TPI nur im zweiten Stamm aktiv waren. Genauer betrachtet wird diese Auflösung wie folgt erreicht:

Der erste Stamm konzentrierte sich auf die Synthese von pep, wodurch Teile der oberen Glycolyse umgangen wurden. Stattdessen nutzte der erste Stamm die Fructose-6-Phosphat-Aldolase (F6PA)-Reaktion, die unter anderem f6p in dha und g3p umwandelt. g3p wurde durch GAPD weiter umgewandelt, und dha wurde sezerniert und vom zweiten Stamm aufgenommen und weiterverarbeitet, wie im nächsten Punkt erläutert wird. Durch die Reaktionen, die im ersten Stamm der Gemeinschaft aktiv waren, konnte eine höhere Konzentration von f6p als im Einzelstamm erreicht werden, sodass die thermodynamische Triebkraft von GAPD in der Gemeinschaftslösung erhöht werden konnte. Um diese höhere Konzentration zu erreichen, musste auch die Konzentration von Glucose-6-phosphat (g6p) für die resultierende f6p-synthetisierende Glucose-6-Phosphat-Isomerase-Reaktion (PGI; $\Delta_r G'^{\circ}$ von +7,01 kJ/mol) erhöht werden.

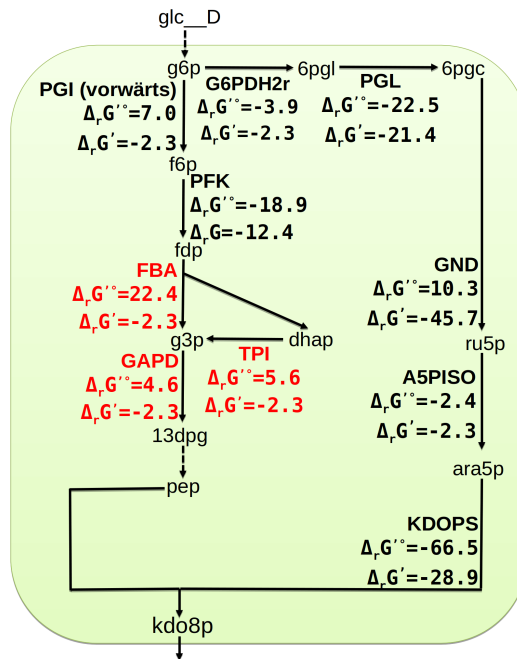
Der zweite Stamm konzentrierte sich auf die Produktion von 6pgc aus dem vom ersten Stamm aufgenommenen dha. Zunächst wurde dabei die reverse F6PA ($\Delta_r G'^{\circ}$ von -15,85 kJ/mol) verwendet, um f6p aus dha sowie g3p herzustellen. Das g3p wurde aus einem Teil des f6p unter Verwendung der g3p-synthetisierenden Reaktionen TPI sowie FBA zurückgewonnen. Durch den negativen $\Delta_r G'^{\circ}$ der inversen F6PA konnten, im Vergleich zum ersten Stamm, im zweiten Stamm niedrigere Konzentrationen von dha sowie g3p eingesetzt werden. Dies führt zu einer höheren thermodynamischen Triebkraft für TPI sowie FBA. Die verbleibende Menge an f6p wurde für die Synthese von 6pgc über die Reaktionen der reversen PGI, G6PDH2r sowie PGL verwendet. Um eine höhere thermodynamische Triebkraft in all diesen Reaktionen zu erreichen, waren die Konzentrationen von dha, f6p, g6p und, wie bereits erwähnt, g3p im Vergleich zum ersten Stamm reduziert. Die Reaktionen zur Synthese von 6pgc aus f6p konnten dennoch aufgrund ihres negativen $\Delta_r G'^{\circ}$ weiterhin ablaufen. (3) Das vom zweiten Stamm synthetisierte 6pgc wurde vom zweiten Stamm sezerniert und vom ersten Stamm aufgenommen. Das 6pgc wurde dann in zwei Stoffwechselwegen weiterverarbeitet, die der Produktion von pep sowie ara5p und damit den Ausgangsstoffen für kdo8p dienen. Dabei wurde pep über g3p sowie Pyruvat (pyr) über den Entner-Doudoroff-Weg produziert. Dieser beinhaltet u.a. die 6-Phosphogluconat-Dehydratase-Reaktion (EDD), die einen sehr negativen $\Delta_r G'^{\circ}$ von -42,8 kJ/mol aufweist. Letzteres ermöglichte eine höhere Variabilität bei der Einstellung der betroffenen Metabolitkonzentrationen, was wiederum der MDF-Optimierung half. Neben diesem Weg wurde g3p auch durch glycolytische Reaktionen wie F6PA gebildet. Das auf diesen Wegen erzeugte g3p wurde dann über weitere glycolytische Reaktionen wie GAPD zu pep weiterverarbeitet. Der andere Ausgangsstoff von kdo8p, ara5p, wurde über die beiden Reaktionen Phosphogluconat-Dehydrogenase (GND) und Arabinose-5-Phosphat-Isomerase (A5PISO) gebildet. Der einschränkende Effekt des hohen $\Delta_r G'^{\circ}$ von GND auf die netzwerkweite thermodynamische Triebkraft wurde teilweise durch die Negativität des $\Delta_r G'^{\circ}$ der anderen Reaktionen abgeschwächt, welche dadurch variablere Metabolitkonzentrationen erlauben. Schließlich wurde ara5p und pep durch die Reaktion der 3-Deoxy-D-mannoctulosonat-8-phosphat-Synthase (KDOPS, von -66,51 kJ/mol) zu kdo8p umgewandelt. Dieses wurde dann durch eine Transportreaktion in die Umwelt sezerniert. Neben den genannten aktiven Reaktionen dieser Lösung fanden sich weitere aktive Reaktionen. Diese dienten u.a. dem stöchiometrischen Ausgleich

zentraler Cofaktorraare wie ATP und ADP sowie NADP und NADPH. Da diese Reaktionen jedoch nicht beschränkend wirkten, werden sie nicht weiter beschrieben.

Zusammenfassend lässt sich sagen, dass der MDF-Vorteil der Gemeinschaft in der betrachteten Einzellösung auf die gleiche Weise erreicht wurde wie im kinetischen Beispielmodell dieses Kapitels: Durch eine geeignete Aufteilung des Weges von Glucose zu kdo8p in den beiden Stämmen konnten die MDF-limitierenden Metabolitkonzentrationen, hier dha, f6p, g3p sowie g6p (Abb. 5.5), in jedem Stamm unterschiedlich eingestellt werden. Auf diese Weise wurde ein insgesamt höherer MDF erreicht. Dieses Grundprinzip ließ sich zudem durch eine Konzentrationsvariabilitätsanalyse (Algorithmus OptMDF5) der kritischen Metabolite (Tab. 5.5) belegen. Diese zeigte zum einen, dass die Konzentrationen der kritischen Metabolite im ersten Stamm immer höher waren als im zweiten Stamm, was die Konzentrationsanpassungen belegt. Andererseits zeigt sie, dass sich die minimalen und maximalen Konzentrationen der kritischen Metabolite nicht überlappten, was bedeutet, dass die Gemeinschaftslösung nicht mit nur einem Stamm erreicht werden kann.

Erwartungsgemäß hat die Gemeinschaft außerdem in der betrachteten Lösung neue thermodynamische Flaschenhälse. Diese waren u.a. das bereits bekannte Dreieck aus FBA, TPI und GAPD. Auch wenn eine Gemeinschaft mit mehr als zwei Stämmen diese Flaschenhälse auflösen könnte, wurde mit *ecolicore2triple* kein MDF-Vorteil festgestellt.

Ausschnitt der Einzelspezies-Lösung
 OptMDF ≈ 2.3 kJ/mol



Ausschnitt der Gemeinschafts-Lösung
 OptMDF ≈ 2.9 kJ/mol

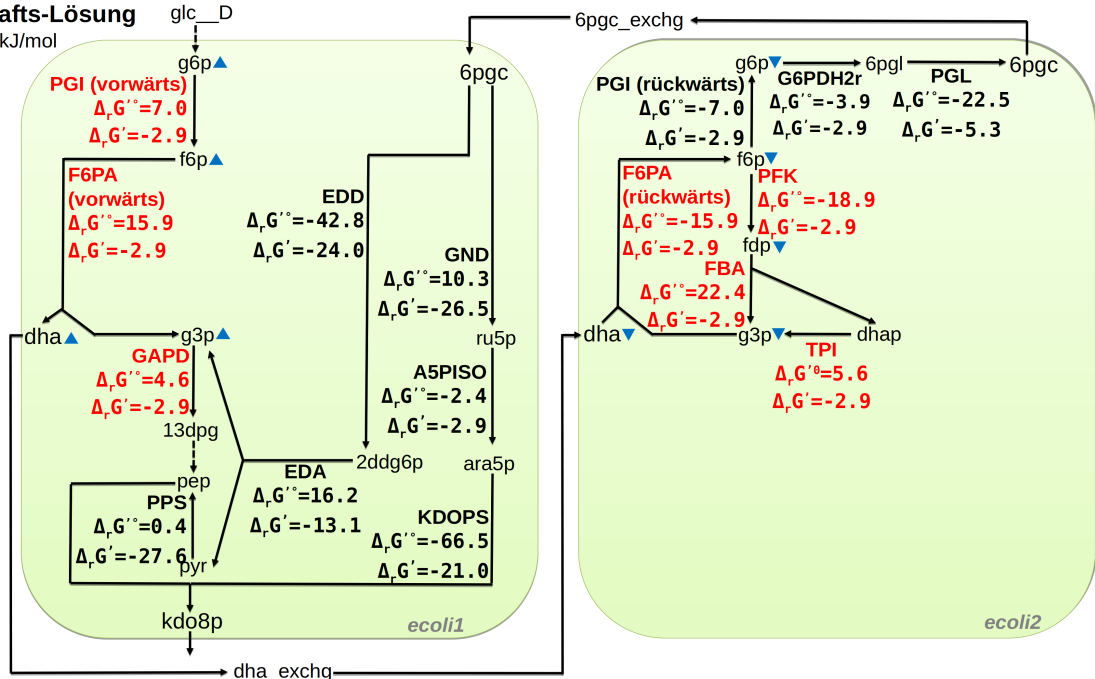


Abbildung 5.5: Ausschnitt zentraler Reaktionen der im Text erläuterten MDF-optimale Einzelspezies- und Gemeinschafts-Lösung für die Synthese von *kdo8p*. Schwarze Pfeile stehen für aktive Reaktionen, gestrichelte Pfeile für mehrere hintereinander ablaufende Reaktionen und blaue Dreiecke dafür, ob eine Metabolitkonzentration in einem Stamm der Gemeinschaft gegenüber der Einzellösung erhöht oder erniedrigt ist. Die gezeigten $\Delta_r G'$ und $\Delta_r G'^{\circ}$ haben die Einheit kJ/mol. Die $\Delta_r G'$ wurden aus der spezifischen hier dargestellten und mit ASTHERISC errechneten MDF-optimale Lösung entnommen. Rote $\Delta_r G'$ -Werte deuten thermodynamische Flaschenhälse an, also alle Reaktionen deren möglicher $\Delta_r G'$ -Bereich auf den MDF fixiert ist. Schwarze $\Delta_r G'$ -Werte weisen eine MDF-Variabilität in der MDF-optimale Lösung auf. Alle Reaktions- und Metabolit-Identifizierer sind der BiGG-Datenbank entnommen [77].

Tabelle 5.5: Mittels Konzentrations-Variabilitätsanalyse ermittelte Konzentrationsbereiche MDF-kritischer Metabolite in den Gemeinschafts-Kompartimenten der dargelegten Beispiellösung zur Produktion von *kdo8p*, unter Forderung des maximalen MDF der Gemeinschaft.

	ecoli1		ecoli2	
	Minimum [mM]	Maximum [mM]	Minimum [mM]	Maximum [mM]
dha_c	0,59826	0,59905	0,058969	0,059048
f6p_c	0,37078	0,37128	0,001	0,0010027
g3p_c	0,32482	0,32525	0,090047	0,090286
g6p_c	19,973	20	0,0019629	0,053238

5.6 Diskussion

Das Hauptergebnis dieses Kapitels ist ASTHERISC, ein neuer Algorithmus zur Identifizierung mikrobieller Gemeinschaften, vornehmlich aus mehreren Stämmen einer Spezies, welche die netzwerkweite thermodynamische Triebkraft einer Produktsynthese im Vergleich zur Nutzung eines Stammes erhöhen. Der zentrale methodische Ansatz von ASTHERISC besteht darin, bestimmte Abschnitte des Produktweges über verschiedene Stämme aufzuteilen und einen Austausch von Zwischenprodukten zwischen diesen Stämmen zu ermöglichen. Dadurch wird eine höhere thermodynamische Triebkraft für den gesamten Produktsyntheseweg durch die beteiligten Stämme erreicht, die über das hinausgeht, was ein einzelner Stamm erreichen kann. Während es bei den aktiven Reaktionen der verschiedenen Stämme zu Überschneidungen kommen kann, führt die Verwendung von mehr als einem Stamm zu zusätzlichen Freiheitsgraden hinsichtlich der erreichbaren MDF. Dies ist auf die potentielle Variation der Metabolitkonzentrationen in den verschiedenen Stämmen zurückzuführen. Durch die Verwendung mehrerer Stämme kann auf diese Weise potentielle thermodynamische Flaschenhälse umgangen werden, die entstehen, wenn eine Reaktion einen Metaboliten mit hoher Konzentration erfordert, während eine andere Reaktion auf denselben Metaboliten mit niedriger Konzentration angewiesen ist.

In den letzten Jahren wurde mikrobiellen Gemeinschaften und ihrem Potential für biotechnologische Produktionsprozesse große Aufmerksamkeit gewidmet. Trotz des Aufbaus und der Analyse verschiedener Einzelstamm-Gemeinschaften, die sich insbesondere durch wesentliche metabolische Abhängigkeiten zwischen den Stämmen auszeichnen, mangelt es auf diesem Gebiet nach wie vor an automatisierten Methoden für das Design von Gemeinschaften mit erwünschten Eigenschaften [51]. Wichtige Beispiele sind dabei FLYCOP [138] und DOLMN [139]. Mit ASTHERISC wird nun darin eine neue Kategorie von Algorithmen eingeführt, die auf der Grundlage thermodynamischer Prinzipien arbeitet. Der DOLMN-Ansatz, der eine optimale Aufteilung des Stoffwechsels einer Art bei gleichzeitiger Begrenzung der Reaktionen pro Stamm anstrebt, weist gewisse Parallelen zu ASTHERISC auf. Jedoch werden in DOLMN keine thermodynamischen Nebenbedingungen verwendet, was, soweit dem Autor bekannt, ein Novum von ASTHERISC im Bereich der Modellierung mikrobieller Gemeinschaften darstellt.

Die Anwendung von ASTHERISC bei der Untersuchung an zwei reduzierten Modellen und einem genomskaligen Modell von *E. coli* identifizierte eine hohe Anzahl an Metaboliten, bei deren Synthese eine *E. coli*-Gemeinschaft aus mehreren Stämmen in der Tat erhöhte netzwerkweite thermodynamische Triebkräfte im Vergleich zu einem Einzelstamm erzeugen könnte. Diese Gemeinschaften könnten einen biotechnologischen Weg darstellen, um die Synthese spezifischer Metabolite in ihrer Rate zu erhöhen oder überhaupt erst thermodynamisch zu ermöglichen. ASTHERISC bietet in Verbindung mit den entwickelten Programmpaketen hierzu zudem einen automatisierten Weg zur Identifizierung solcher vorteilhaften Multi-Stamm-Gemeinschaften.

Die mit ASTHERISC ermittelten Lösungen belegen zwar das Potential für die Erzeugung erhöhter thermodynamischer Triebkräfte durch die vorhergesagten Multi-Stamm-Gemeinschaften, doch eine praktische Umsetzung würde einen nicht unerheblichen Aufwand erfordern. In erster Linie wird das Design dieser Stämme die Deaktivierung bestimmter Stoffwechselreaktionen (z.B. durch Gen-Knockouts) in Übereinstimmung mit der identifizierten Gemeinschaftslösung erfordern. Die Formulierung geeigneter Knockout-Strategien, die nur die relevanten Abschnitte der Stoffwechselwege in den jeweiligen Stämmen aktiv halten, kann durch spezielle Algorithmen für das Stammdesign berechnet werden, z.B. durch solche, die auf Minimal Cut Sets basieren (Gleichungen (2.32) bis (2.33)). Die praktische Umsetzung dieser gezielten Knockout-Strategien könnte jedoch zu instabilen Stämmen führen oder sogar das Wachstum komplett behindern. Auch wenn die (stöchiometrische) Durchführbarkeit von Wachstum bei der Berechnung der entsprechenden MCS festgesetzt werden kann, könnte sich ein zweistufiger Ansatz als pragmatischer erweisen. In diesem Szenario wird den Stämmen zunächst erlaubt, zu wachsen, bevor sie in eine spezielle Produktionsphase übergehen (z.B. über dynamische Stoffwechselschalter [140, 141]). Dieser Ansatz umgeht auch potentielle Einschränkungen im Zusammenhang mit dem MDF des Weges der Produktsynthese, insbesondere wenn Wachstum und Produktion miteinander verflochten sind. Dies ähnelt den Berechnungen in diesem Kapitel, die auf die Optimierung der Aufteilung der Produkt-Synthesewege abzielten und Überlegungen zur Durchführbarkeit des Wachstums außer Acht ließen.

Im Rahmen der praktischen Umsetzung einer von ASTHERISC vorhergesagten Gemeinschaft muss außerdem der Transport von Austauschmetaboliten zwischen den Stämmen erleichtert werden (z.B. durch Überexpression spezieller Transportsysteme). Diese Transportsysteme sind jedoch nur für bestimmte Metabolite verfügbar oder bekannt, sodass ASTHERISC-Lösungen mit einer begrenzten Zahl an Austauschmetaboliten günstiger erscheinen. Dennoch ist es erwähnenswert, dass viele der identifizierten Gemeinschaftslösungen für verschiedene Produkte oft ähnliche Mengen an Austauschmetaboliten verwenden. Folglich könnte nach der Entdeckung oder Implementierung geeigneter Transportermechanismen für diese wichtigen Metabolite eine breitere Palette von Stammdesigns möglich werden. Bei den Überlegungen zum Metabolittransport wurde davon ausgegangen, dass für alle Austauschreaktionen ein Konzentrationsgefälle zwischen den zellulären und externen Kompartimenten bestehen muss. Die mit dem Transport verbundenen Kosten, wie z.B. der Energieverbrauch für die aktive Bewegung eines Metaboliten durch die Membran, wurden jedoch nicht ausdrücklich berücksichtigt (abgesehen von den bereits im Basismodell enthaltenen Transportern dieser Art). Diese zusätzlichen Kosten könnten sich auf die Produktausbeute auswirken und sogar den MDF-Vorteil einer Gemeinschaftslösung zunichtemachen. Um die

Sensitivität dieser Lösungen bei der Einführung von Energiekosten für den Metabolitransport zu bewerten, wurden alle Berechnungen unter Verwendung einer modifizierten Version des `ecolicore2double`-Modells wiederholt. In dieser Variante verbraucht jeder eingeführte (zusätzliche) Metabolitaustausch $1/3$ ATP pro transportiertem Metabolitmolekül. Während es nur wenige spezifische Änderungen in der Gruppe der Metabolite gibt, bei denen ein MDF-Vorteil gefunden werden kann, gab es beim Gesamtanteil der Metabolite, die einen MDF-Vorteil innerhalb einer Gemeinschaftslösung aufweisen, keine signifikanten Änderungen. Interessanterweise würde in der Einzelstammlösung der Export des Endproduktes aufgrund dieser Modifikation auch ATP erfordern, was möglicherweise die Struktur des dort aktiven Reaktionsnetzwerks verändert und eine Gemeinschaftslösung begünstigt. Die entsprechenden Skripte und Ergebnisdateien sind ebenfalls in den schon erwähnten GitHub-Repositories von `CommModelPy` und dem `ASTHERISC`-Paket verfügbar.

Eine weitere wichtige Überlegung zur Realisierung eines berechneten Gemeinschaftsdesigns ist die langfristige Stabilität der Gemeinschaft. Dies könnte die Anwendung genetischer Methoden und damit realisierter allgemeiner Prinzipien zur Stabilität von Gemeinschaften erforderlich machen, wie sie in [18] und kürzlich in [58] entwickelt wurden. Um die berechneten maximalen thermodynamischen Triebkräfte für die Gemeinschaften zu erreichen, müssen zudem die Metabolitkonzentrationen in die entsprechenden Bereiche fallen, die für einen optimalen MDF erforderlich sind. Es ist jedoch nicht von vornherein klar, ob diese Bedingung erfüllt ist oder ob durch geeignete Selektionsmechanismen Anpassungen durch adaptive Evolution vorgenommen werden können. Der Nachweis, dass die berechneten Gemeinschaften eine ökonomisch signifikante Verbesserung bieten können, muss daher noch erbracht werden. Dennoch ist es wichtig zu erkennen, dass viele der hier diskutierten Herausforderungen für die Implementierung jeder (berechneten) synthetischen Gemeinschaft relevant sind.

Die `ASTHERISC`-Läufe in den Berechnungen dieses Kapitels liefen weitestgehend stabil. Potentielle numerische Probleme, die etwa durch Präzisionstoleranzen des genutzten MILP-Lösers entstehen würden, werden zum Teil direkt durch Löser-Einstellungen verhindert oder von der `ASTHERISC`-Implementierung erkannt und umgangen. So werden beispielsweise vom Löser gelieferte „Gemeinschafts-Scheinlösungen“, bei denen nur ein Stamm effektiv genutzt wird, erkannt und verworfen. Es ist dabei wichtig zu erwähnen, dass die gesetzten Grenzen der Rechenzeit die Ergebnisse beeinflussen können. Dies zeigt sich etwa in den geringeren Prozentsätzen des MDF-Vorteils bei `ecolicore2triple` im Vergleich zu `ecolicore2double` (Tab. 5.2). Der MDF-Vorteil in `ecolicore2triple` sollte mindestens so hoch sein wie in `ecolicore2double`, da ersteres implizit alle Zwei-Stamm-Lösungen einschließt. Dieser Effekt ist auf die Wahrscheinlichkeit zurückzuführen, dass die MILP-Zeitgrenzen im größeren Modell häufiger erreicht werden. Dennoch zeigten die Durchläufe des `ecolicore2triple`-Gemeinschaftsmodells, dass eine größere Anzahl von Stämmen den Gesamtvorteil der Gemeinschaft erhöhen könnte. Zu allem zeigten die mit dem `iML1515double`-Modell durchgeführten Berechnungen, dass die Implementierung des `ASTHERISC`-Algorithmus mit einer angemessenen Laufzeit auf genomskalige Modelle angewendet werden kann. Der Vergleich der ermittelten Lösungen für `ecolicore2double` und `iML1515double` deutet darauf hin, dass für bestimmte Metabolite alternative MDF-optimierende Pfade in einem genomskaligen

Modell gefunden werden könnten. Diese Wege könnten jedoch weniger verbreitete Routen (mit potentiell geringerer Kapazität) beinhalten als die im reduzierten Modell gefundenen Lösungen.

Die mit ASTHERISC erzielten Ergebnisse hängen auch von den verwendeten Einstellungen ab (z.B. ein minimal geforderter MDF-Vorteil und eine minimale Produktausbeute), und die Berechnungen beruhen zum Teil auf unsicheren Parametern oder Annahmen, wie z.B. die Konzentrationsbereiche der Metabolite oder die mit dem eQuilibrator [45] berechneten $\Delta_r G^\circ$. Anders als etwa in TCOSA (siehe vorheriges Kapitel) wurde zudem keine Behandlung für das Wachstum relevanter thermodynamischer Flaschenhälse (durch Abschwächung von deren $\Delta_r G^\circ$) durchgeführt, die den Lösungsraum in den Analysen mit ASTHERISC verändern könnte. Alle Berechnungen können jedoch leicht mit alternativen Werten wiederholt werden, wenn diese als relevanter oder realistischer erachtet werden. Darüber hinaus sind die entwickelten Pakete CommModelPy und ASTHERISC programmtechnisch unabhängig von den verwendeten Modellen, sodass sie direkt für die Erstellung und Untersuchung anderer Gemeinschaftsmodelle eingesetzt werden können. Obwohl noch nicht damit angewendet, könnte ASTHERISC auch bei Multi-Spezies-Gemeinschaften eingesetzt werden, um z.B. zu untersuchen, welcher Produktsyntheseweg (zusammen mit den damit verbundenen Metabolitaustauschen) innerhalb einer Multi-Spezies-Gemeinschaft zum höchsten MDF führt, und ob dieser MDF einem potentiell vorhandenen Einzelspezies-Weg überlegen ist. Schließlich können auch nützliche Anwendungen der ASTHERISC-Methode zur Entdeckung geeigneter Kompartimentierungsstrategien erfolgen, die die thermodynamischen Triebkräfte in zellfreien Produktionssystemen maximieren.

Beim Design von Gemeinschaften durch ASTHERISC wird die Tatsache ausgenutzt, dass mehrere Stämme die Anpassung unterschiedlicher kritischer Metabolitkonzentrationen in den verschiedenen Stämmen ermöglichen. Während ASTHERISC diesen Freiheitsgrad zur Maximierung der netzwerkweiten thermodynamischen Triebkraft nutzt, könnten künftige Arbeiten darauf abzielen, die Produktsyntheserate in der Gemeinschaft direkt zu maximieren, indem eine Aufteilung der Stoffwechselwege (wiederum mit spezifischen Metabolitkonzentrationen in den einzelnen Stämmen) gefunden wird, die unter Berücksichtigung der kinetischen Geschwindigkeitsgesetze den Gesamtproduktfluss optimiert. Idealerweise könnte man, wenn ein kinetisches Modell des (zentralen) Stoffwechsels verfügbar wäre, direkt nach einer solchen optimalen Lösung suchen. Das exemplarische kleine kinetische Modell dieser Arbeit (Abb. 5.2), mit dem praktisch gezeigt wurde, dass ASTHERISC eine plausible Heuristik zur Bestimmung des gemeinschaftlichen Nutzens von Produktsyntheseraten darstellt, ist ein erster Ansatz in diese Richtung.

6 Zusammenfassende Diskussion und Ausblick

In dieser Arbeit wurden folgende drei neue Methodiken und Ansätze im Gebiet der constraint-basierten Modellierung vorgestellt:

1. Mit der entwickelten sMOMENT-Methode lassen sich enzymatische Nebenbedingungen bei minimaler Vergrößerung des zugrundeliegenden Modells integrieren, wobei die Resultate mit typischen Optimierungen im Vergleich zu den vorherigen Methoden MOMENT sowie GECKO identisch bleiben. Die sMOMENT-Implementierung AutoPACMEN integriert, als erstes Programmpaket, die benutzerfreundliche Akquirierung aller relevanten Parameter, die Erstellung eines constraint-basierten Modells mit enzymatischen Nebenbedingungen und die anschließende Kalibrierung der in das Modell integrierten enzymatischen Parameter. Als erste Anwendung von AutoPACMEN wurde das sMOMENT-basierte Modell *iJO1366** erstellt. Dies stellt gleichzeitig das erste Modell von *E. coli* mit enzymatischen Nebenbedingungen dar, das ein plausibles Sekretionsverhalten sowohl unter aeroben als auch anaeroben Bedingungen und bei unterschiedlichen Wachstumsraten aufweisen kann. Vergleichende Berechnungen von MCS mit diesem Modell und dessen Vorgängermodell *iJO1366* zeigten zudem, dass die Einbindung enzymatischer Nebenbedingungen tatsächlich zu gänzlich neuen Vorhersagen gentechnischer Interventionen führen kann.
2. Mit dem TCOSA-Framework können die Auswirkungen der Änderung von Reaktionsspezifitäten bezüglich Cofaktoren im Hinblick auf die netzwerkweite thermodynamische Triebkraft analysiert werden. Durch die Anwendung von TCOSA auf ein Modell von *E. coli* und dessen NAD(P)(H)-Spezifitäten konnte zum ersten Mal gezeigt werden, dass diese Spezifitäten offensichtlich stark durch einen evolutionären Antrieb zur Maximierung der netzwerkweiten thermodynamischen Triebkraft beeinflusst wurden. Darüber hinaus wurde, anders als etwa in [104], gezeigt, dass ein hypothetisches Netzwerk mit nur einem Redoxcofaktorpaar thermodynamisch deutlich ungünstiger wäre. Im Gegensatz dazu wurde gezeigt, dass ein Netzwerk mit drei Cofaktoren nur bei einem deutlich niedrigeren Standard-Redoxpotential des hypothetischen dritten Cofaktorpaares nennenswerte Vorteile bringen würde. Aus den Überlegungen zu TCOSA und seiner Anwendung könnte auch ein mögliches Designprinzip für die thermodynamisch optimierte biotechnologische Synthese eines Produktes entwickelt werden. Nach diesem Prinzip kann nach Cofaktor-Austauschen gesucht werden, welche die netzwerkweite thermodynamische Triebkraft bei der Synthese eines Produktes optimieren.
3. Mit der ASTHERISC-Methode ist es möglich, Produktsynthesen zu identifizieren, bei denen eine Gemeinschaft mehrerer Stämme einen Vorteil gegenüber einem einzelnen Stamm hat. Dieser Vorteil bezieht sich dabei auf die netzwerkweite thermodynamische Triebkraft, die bei der Produktsynthese auftritt. Die Relevanz eines solchen Konzepts für das Design einer Gemeinschaft und die Erkenntnis, dass das weithin akzeptierte Konzept der Arbeitsteilung vage ist, wurde zuvor in einem hypothetischen Stoffwechselweg gezeigt. Insbesondere wurde festgestellt, dass sich unterscheidende (stammabhängige) Metabolitkonzentrationen angenommen werden müssen, damit sich ein Flussvorteil einer Gemeinschaft manifestieren kann. Ein anschauliches kinetisches Modell zeigte auch, dass die Vorteile einer Gemeinschaft zumindest für weite Bereiche der kinetischen Parameter auf der Grundlage

der Netzwerkstruktur und der Identifizierung kritischer Metabolitkonzentrationen vorhergesagt werden können. Diese praktische Demonstration unterstreicht, dass ASTHERISC eine biologisch plausible Heuristik zur Bestimmung des Nutzens einer Gemeinschaft im Sinne eines thermodynamischen Vorteils bei einer Produktsynthese ist. ASTHERISC hat somit eine auf der Thermodynamik basierende und vor allem berechenbare Methode eingeführt, um bestimmen zu können, wann und warum der Einsatz von Gemeinschaften thermodynamisch vorteilhaft sein könnte. Dies wurde auch durch Anwendungen auf reduzierte sowie genomskalige Modelle von *E. coli* demonstriert, wobei dutzende potentieller thermodynamischer Gemeinschaftsvorteile bei der Synthese von Produkten identifiziert wurden.

Die drei Methoden stellen verbesserte Modellierungstechniken dar. Jeder dieser neuen Ansätze birgt zudem ein Potential für ein verbessertes rationales Design biotechnologischer Prozesse, an denen ein oder mehrere biologische Stämme beteiligt sind. sMOMENT erleichtert durch die minimale Erweiterung eines Modells mit enzymatischen Nebenbedingungen die erfolgreiche Anwendung rechenintensiver Verfahren, wie z.B. MCS. Sowohl ASTHERISC als auch TCOSA, die jeweils von unterschiedlichen Designprinzipien geleitet werden, eröffnen potentiell neue Wege für thermodynamisch basierte Möglichkeiten biotechnologischer Anwendungen.

Trotz der Erweiterung der constraint-basierten Modellierungstechniken durch die drei in dieser Arbeit vorgestellten neuen Methoden sind diese nicht frei von potentiellen Nachteilen, wie z.B. der Unsicherheit bei den genutzten Daten. So sind z.B. sowohl die enzymatischen als auch die thermodynamischen Nebenbedingungen, die hier verwendet werden, mit erheblichen Unsicherheiten in den zugrundeliegenden Daten behaftet. Im Falle von sMOMENT betrifft dies vor allem die Unsicherheiten bezüglich der k_{cat} -Werte und des Gesamt-Enzym-pools. Trotz der Verwendung eines komplexeren Algorithmus zur Ableitung plausibler k_{cat} -Werte war, wie bereits erläutert, eine umfangreiche Kalibrierung der enzymatischen Parameter notwendig, um glaubwürdige Ergebnisse zu erhalten. Inwieweit die in Datenbanken dokumentierten k_{cat} -Werte für Modelle, die verschiedene Bedingungen untersuchen, relevant sind und ob diese Werte tatsächlich eher „scheinbare“ k_{app} -Werte für eine bestimmte Bedingung darstellen, bleibt eine ungelöste Frage. Die biologisch unpräzise Definition einer durchschnittlichen Sättigung aller Enzyme als Teil der Enzym-poolbestimmung ist, wie bereits erläutert, eine Vereinfachung, die eine weitere Kalibrierung der k_{app} -Werte für das Gleichgewicht erforderlich macht. Wie jedoch anhand von *iJO1366** gezeigt werden konnte, ist letzteres durchaus praktikabel, zumindest für die Vorhersage bestimmter Aspekte des Stoffwechsels auch unter variierenden Bedingungen. Die für TCOSA und ASTHERISC verwendeten Konzentrationsbereiche der Metabolite und die offensichtlich falsch berechneten bzw. fehlenden $\Delta_r G^\circ$ -Werte stellen erhebliche Unsicherheitsfaktoren für thermodynamische Methoden der constraint-basierten Modellierung dar. Obwohl die Konzentrationsbereiche der Metabolite durch Metabolom-Messungen eingegrenzt werden könnten, könnten diese Werte stark von den Umweltbedingungen der jeweiligen Experimente abhängen. Umgekehrt stellt sich bei den $\Delta_r G^\circ$ -Werten das Problem der Abhängigkeit von Umweltbedingungen nicht. Aus rechnerischer Sicht könnte die Entwicklung alternativer Algorithmen, wie z.B. der dGPredictor zur Vorhersage von $\Delta_r G^\circ$ -Werten [142], eine Möglichkeit darstellen, die Fehlerquote zu reduzieren. In diesem Zusammenhang könnte auch die Einbeziehung

von künstlicher Intelligenz in Form von maschinellem Lernen ein plausibler Weg sein, die Unsicherheit bei den Daten zu verringern. Ansätze dieser Art sind bereits für die Vorhersage von Enzymkinetikparametern bekannt (z.B. [143]), bleiben aber derzeit weiterhin anfällig für potentiell erhebliche Unsicherheiten.

Eine experimentelle Validierung der in dieser Arbeit entwickelten neuen Designprinzipien steht derzeit noch aus. Bezüglich der verwendeten grundlegenden Prinzipien der Thermodynamik [2] und der Michaelis-Menten-Kinetik [1] ist eine umfassende Bewertung, inwieweit ihre vereinfachte Integration in den hier verwendeten Methoden korrekt und relevant sein könnte, noch nicht abgeschlossen. Eine weitere wichtige Überlegung ist die Rolle von kinetischen Faktoren, die bei der constraint-basierten Modellierung nicht berücksichtigt werden. Insbesondere die Auslassung des Sättigungsterms in der zerlegten reversiblen Michaelis-Menten-Kinetik (Gleichung (2.5)) lässt eine Vielzahl an Möglichkeiten zur Beschränkung und Beeinflussung von Metabolitkonzentrationen und Reaktionsflüssen außer Acht. Eine effektiv nutzbare Linearisierung dieser zusätzlichen kinetischen Terme, die für eine allgemeine, auf Restriktionen basierende Modellierung und praktische Anwendung unerlässlich sind, sowie der bisher nicht berücksichtigten Komponenten des thermodynamischen Terms in der zerlegten reversiblen Michaelis-Menten-Kinetik (Gleichung (2.4)) steht jedoch noch aus. Gleichzeitig mangelt es auch an einer Betrachtung sowie Integration von Reaktionskinetiken, die nicht der Michaelis-Menten-Kinetik folgen. Unabhängig von der spezifischen Reaktionskinetik bleiben die Grundregeln der Thermodynamik bestehen, darunter primär die Notwendigkeit einer positiven thermodynamischen Triebkraft für den Reaktionsablauf. Allerdings führen zelluläre Mechanismen, wie etwa das bereits erwähnte Metabolit-Kanalieren [111], eine zusätzliche Unsicherheitsebene bei der Bestimmung der thermodynamisch möglichen Zustände ein.

Die Unsicherheit der Resultate dieser Arbeit, die sowohl von der Datenseite als auch von der praktischen Anwendbarkeit herrührt, weist auch auf grundsätzliche Vor- und Nachteile der hier entwickelten constraint-basierten Modellierungstechniken im Vergleich zu den bereits erwähnten verwandten Konzepten der RBA [95] und der ME-Modelle [96] hin. Letztere Ansätze sind zwar in der Lage, verschiedene Arten an Daten effektiv zu integrieren, sind aber gleichzeitig entsprechend anfälliger für Datenunsicherheiten. Insbesondere bei schlecht charakterisierten Organismen ist ein constraint-basierter Ansatz, der einen Fokus auf zuverlässigere Daten ermöglicht, von Vorteil. Die Integration von Translations- und Transkriptionsprozessen in die oben genannten Ansätze könnte dagegen präzisere Vorhersagen ermöglichen. Diese Verbesserung geht jedoch zusätzlich zulasten der Rechenintensität, sodass Berechnungen mit besonders großen genomskaligen Netzwerken nur mit effizient formulierten Techniken der constraint-basierten Modellierung möglich sind. Wie sMOMENT hierzu zeigt, gibt es selbst in etablierten Methodiken weiterhin Potentiale, die zugrundeliegende Modellkomplexität weiter zu rationalisieren und damit die Rechenintensität constraint-basierter Berechnungen zu reduzieren.

Trotz aller beschriebenen Unsicherheiten könnte insbesondere TCOSA über das Gebiet des Metabolic Engineerings auch im Bereich der Evolutionsbiologie nützliche Beiträge leisten. Das mit TCOSA vorgestellte Anwendungsbeispiel konzentriert sich dabei auf die Cofaktorpaare NAD(H) sowie NADP(H), und zeigt eine wünschenswerte evolutionäre Anpassung der Redoxcofaktor-

Spezifitäten zur Optimierung der netzwerkweiten thermodynamischen Triebkraft. Abstrakter betrachtet kann das Vorhandensein redundanter Cofaktorpaare als eine Art „metabolische Kompartimentalisierung“ verstanden werden. Anders als etwa in ASTHERISC, wo ein thermodynamischer Vorteil einer Produktsynthese durch mehrere Kompartimente erfolgte, genügt hierbei ein Kompartiment mit mehreren Metaboliten, welche verschiedene physiologische Rollen übernehmen. TCOSA kann somit einen neuen Ansatz für die constraint-basierten Modellierung im Kontext der Evolutionsbiologie darstellen, der die Thermodynamik als wertvolle Ergänzung zu früheren relevanten constraint-basierten Methoden, wie etwa [30–32], beinhaltet.

Neben der bereits erwähnten möglichen experimentellen Validierung der in dieser Arbeit vorgestellten Ergebnisse und Prinzipien könnte als nächster Schritt eine Kombination oder weitergehende Anwendung der hier vorgestellten Methoden in Betracht gezogen werden. Im Hinblick auf Kombinationen ist beispielsweise die gleichzeitige Anwendung enzymatischer und thermodynamischer Nebenbedingungen denkbar, wie sie bereits mit anderen Methoden dieser Kategorien praktiziert wurde [144]. Dies erfordert jedoch eine komplexere gleichzeitige Berücksichtigung von Unsicherheiten, die nun sowohl mit enzymatischen als auch mit thermodynamischen Daten verbunden wären, um ein Modell sinnvoll zu verstehen und zu kalibrieren. Insbesondere erfordert in einem solchen Modell die Bestimmung, welche dieser interagierenden Nebenbedingungen eine Grenze für die Optimierung einer Zielfunktion darstellt und ob dies biologisch plausibel ist, ein umfassendes Verständnis sowohl der thermodynamischen als auch der enzymatischen Prinzipien hinter dem Stoffwechsel des modellierten Organismus. Eine weitere Kombinationsmöglichkeit besteht mit den Methoden TCOSA und ASTHERISC, die beide, wie beschrieben, auf ein abstraktes Kompartimentierungskonzept aufbauen. Diese Methoden könnten etwa methodisch kombiniert werden, um thermodynamisch bessere Synthesewege zu ermitteln, wobei sowohl die Gemeinschaft als auch die Cofaktor-Spezifitäten der einzelnen Stämme optimiert werden können. Zudem wurden alle hier vorgestellten Berechnungen ausschließlich mit *E. coli*-Modellen durchgeführt, da *E. coli* als Modellorganismus der Mikrobiologie auch die größte Datenbasis aufweist. Die Anwendung dieser Methoden auf andere Organismen, speziell solche mit deutlich abweichendem Stoffwechsel, könnten jedoch wertvolle Erkenntnisse für die Vorhersage biotechnologisch relevanter Stoffwechselwege und, speziell durch TCOSA, evolutionäre Erklärungen für Teile der Netzwerkstruktur liefern. Im Falle von ASTHERISC, wo im Rahmen dieser Arbeit bewusst nur Stämme einer Spezies berücksichtigt wurden, könnte die Verwendung mehrerer Stämme verschiedener Spezies möglicherweise eine Reihe weiterer thermodynamisch vorteilhafter Synthesewege aufdecken. Die in dieser Arbeit vorgestellten Methoden können daher auch in Zukunft eine wertvolle Rolle bei der Gestaltung nachhaltiger biotechnologischer Prozesse spielen.

7 Anhang

Tabelle A1: Kinetische Gleichungen aller Reaktionen des kinetischen Beispielmodells welches im Rahmen von ASTHERISC entwickelt wurde (Abb. 5.2). In der programmatischen Implementierung des Modells wird jede dieser Gleichungen mit einem eigenen Parameter namens „is_on“ multipliziert. „is_on“ zeigt dabei an ob eine Reaktion aktiv ist (is_on=1) oder nicht (is_on=0) im jeweiligen Kompartiment.

Reaktion	Reaktionskinetik
R ₁	$r_{R_1} = [E]_{R_1} \cdot k_{cat,R_1} \cdot \frac{([S_{ex}]/k_{S,R_1})}{1 + ([S_{ex}]/k_{S_{ex},R_1}) + ([S]/k_{S,R_1})} \cdot \left(1 - e^{\frac{\Delta_r G'_{R_1}}{R \cdot T}}\right)$
R ₂	$r_{R_2} = [E]_{R_2} \cdot k_{cat,R_2} \cdot \frac{([S]/k_{S,R_2})}{1 + ([S]/k_{S_{ex},R_2}) + ([A]/k_{S,R_2}) \cdot ([X]/k_{X,R_2})} \cdot \left(1 - e^{\frac{\Delta_r G'_{R_2}}{R \cdot T}}\right)$
R ₃	$r_{R_3} = [E]_{R_3} \cdot k_{cat,R_3} \cdot \frac{([A]/k_{A,R_3}) \cdot ([M]/k_{M,R_3})}{1 + ([A]/k_{A,R_3}) \cdot ([M]/k_{M,R_3}) + ([B]/k_{B,R_3})} \cdot \left(1 - e^{\frac{\Delta_r G'_{R_3}}{R \cdot T}}\right)$
R ₄	$r_{R_4} = [E]_{R_4} \cdot k_{cat,R_4} \cdot \frac{([C]/k_{C,R_4})}{1 + ([C]/k_{C,R_4}) + ([P]/k_{P,R_4})} \cdot \left(1 - e^{\frac{\Delta_r G'_{R_4}}{R \cdot T}}\right)$
R ₅	$r_{R_5} = [E]_{R_5} \cdot k_{cat,R_5} \cdot \frac{([C]/k_{C,R_5}) \cdot ([X]/k_{X,R_5})}{1 + ([C]/k_{C,R_5}) \cdot ([X]/k_{X,R_5}) + ([P]/k_{P,R_5})} \cdot \left(1 - e^{\frac{\Delta_r G'_{R_5}}{R \cdot T}}\right)$
R ₆	$r_{R_6} = [E]_{R_6} \cdot k_{cat,R_6} \cdot \frac{([P]/k_{S,R_6})}{1 + ([P]/k_{P,R_6}) + ([P_{ex}]/k_{P_{ex},R_6})} \cdot \left(1 - e^{\frac{\Delta_r G'_{R_6}}{R \cdot T}}\right)$
R _{RTB,2}	$r_{RTB,2} = [E]_{RTB} \cdot k_{cat,RTB} \cdot \frac{([B_2]/k_{B_2,RTB,2})}{1 + ([B_2]/k_{B_2,RTB,2}) + ([B_{ex}]/k_{B_{ex},RTB,2})} \cdot \left(1 - e^{\frac{\Delta_r G'_{RTB,2}}{R \cdot T}}\right)$
R _{RTB,3}	$r_{RTB,3} = [E]_{RTB} \cdot k_{cat,RTB} \cdot \frac{([B_{ex}]/k_{B_{ex},RTB,3})}{1 + ([B_3]/k_{B_3,RTB,3}) + ([B_{ex}]/k_{B_{ex},RTB,3})} \cdot \left(1 - e^{\frac{\Delta_r G'_{RTB,3}}{R \cdot T}}\right)$
R ₇	$r_{R_7} = I_{R_7}$
R ₈	$r_{R_8} = k_{M,R_8} \cdot [M]$
R ₉	$r_{R_9} = I_{R_9}$
R ₁₀	$r_{R_9} = k_{M,R_9} \cdot [M]$

Tabelle A2: Differentialgleichungen der Konzentrationsänderungen aller internen Metabolite im kinetischen Beispielmodell, unter Nutzung der Reaktionsdefinitionen in Tabelle A1. Da die externen Metabolite M_{ex} sowie X_{ex} nicht explizit mit einer eigenen Konzentration modelliert und nur implizit als konstante Aufnahme­raten im Modell eingebaut sind, sind sie in dieser Tabelle nicht enthalten. Wo nicht anders angegeben, gelten diese Differentialgleichungen im Einzelstamm und den beiden Stämmen der Gemeinschaft des Modells für die jeweilige Instanz des Metaboliten in jedem einzelnen Kompartiment.

Metabolit	Konzentrationsänderung über Zeit
S_{ex}	$\frac{d[S_{ex}]}{dt} = 0$ (S_{ex} ist konstant auf Konzentration von 4 gesetzt)
S (im Einzelstamm und Stamm 1 der Gemeinschaft)	$\frac{d[S]}{dt} = r_{R1} - r_{R2}$
A (im Stamm 1 der Gemeinschaft)	$\frac{d[A]}{dt} = r_{R2} - r_{R3}$
B (im Einzelstamm)	$\frac{d[B]}{dt} = r_{R3} - r_{R4}$
B (im Stamm 1 der Gemeinschaft)	$\frac{d[B]}{dt} = r_{R3} - r_{RTB,2}$
B (im Stamm 2 der Gemeinschaft)	$\frac{d[B]}{dt} = r_{RTB,3} - r_{R4}$
C (im Einzelstamm und Stamm 2 der Gemeinschaft)	$\frac{d[C]}{dt} = r_{R4} - r_{R5}$
M (im Einzelstamm und Stamm 1 der Gemeinschaft)	$\frac{d[M]}{dt} = r_{R7} - r_{R8} - r_{R3}$
P (im Einzelstamm und Stamm 2 der Gemeinschaft)	$\frac{d[P]}{dt} = r_{R5} - r_{R6}$
P_{ex}	$\frac{d[P_{ex}]}{dt} = 0$ (P_{ex} ist konstant auf Konzentration von 2 gesetzt)
X (im Einzelstamm)	$\frac{d[X]}{dt} = r_{R2} + r_{R10} - r_{R5} - r_{R9}$
X (im Stamm 1 der Gemeinschaft)	$\frac{d[X]}{dt} = r_{R2} + r_{R10} - r_{R9}$
X (im Stamm 2 der Gemeinschaft)	$\frac{d[X]}{dt} = r_{R10} - r_{R5} - r_{R9}$
B_{ex} (zwischen Stamm 1 und 2 der Gemeinschaft)	$\frac{d[X]}{dt} = r_{RTB,2} - r_{RTB,3}$

Tabelle A3: Liste der kinetischen Parameter des im Rahmen von ASTHERISC dargestellten kinetischen Beispielmmodells, mit einer Angabe ob dieser Wert in der Stichprobenanalyse (Tab. 5.2) variiert wird oder nicht.

Parameter	Wert	Wird variiert?
I_{Min}	2,0	Ja
I_{Xin}	0,8	Ja
$k_{cat,R1}$	6	Ja
$k_{cat,R2}$	5	Ja
$k_{cat,R3}$	8	Ja
$k_{cat,R4}$	9	Ja
$k_{cat,R5}$	10	Ja
$k_{cat,R6}$	24	Ja
$k_{cat,RTB}$	1.000.000	Ja
$k_{S_{ex},R1}$	0,18	Ja
$k_{S,R1}$	2,5	Ja
$k_{S,R2}$	0,4	Ja
$k_{A,R2}$	3,1	Ja
$k_{X,R2}$	2,1	Ja
$k_{A,R3}$	4,6	Ja
$k_{M,R3}$	0,4	Ja
$k_{B,R3}$	2,5	Ja
$k_{B,R4}$	0,5	Ja
$k_{C,R4}$	0,9	Ja
$k_{C,R5}$	1,5	Ja
$k_{X,R5}$	2,6	Ja
$k_{P,R5}$	6,8	Ja
$k_{P,R6}$	0,01	Ja
$k_{P_{ex},R6}$	14	Ja
$k_{B,RTB}$	2	Ja
$k_{B_{ex},RTB}$	2	Ja
$k_{M,Mout}$	0,75	Ja
$k_{X,Xout}$	0,5	Ja
[E] (überall)	1	Nein
$\Delta_r G'_{R1}$	0	Nein
$\Delta_r G'_{R2}$	4	Nein

(Fortsetzung von Tab. A3)

$\Delta_r G_{R3}^{\prime\circ}$	-5	Nein
$\Delta_r G_{R4}^{\prime\circ}$	-5	Nein
$\Delta_r G_{R5}^{\prime\circ}$	4	Nein
$\Delta_r G_{R6}^{\prime\circ}$	0	Nein
$\Delta_r G_{RTB}^{\prime\circ}$	0	Nein

Tabelle A4: Liste der standardmäßig mit der Umwelt austauschbaren Metabolite in den Gemeinschaftsmodellen *ecolicore2double*, *ecolicore2triple* sowie *iML1515double*. „_p“ steht für Peri-, „_c“ für das Cytoplasma. Im letzteren Falle wird das Periplasma umgangen.

BiGG-Identifizierer des Metaboliten	Metabolitname	Kann sekretiert werden?	Kann importiert werden?
4crsol_c	P-Kresol	Ja	Nein
5drib_c	5'-Deoxyribose	Ja	Nein
aacald_c	Aminoacetaldehyd	Ja	Nein
ac_p	Acetat	Ja	Nein
amob_c	S-Adenosyl-4-Methylthio-2-Oxobutanoat	Ja	Nein
ca2_p	Calcium 2 ⁺	Nein	Ja
cl_p	Chlorid	Nein	Ja
co2_p	Kohlenstoffdioxid	Ja	Ja
cobalt2_p	Kobalt 2 ⁺	Nein	Ja
cu2_p	Kupfer 2 ⁺	Nein	Ja
etoh_p	Ethanol	Ja	Nein
fe2_p	Eisen 2 ⁺	Nein	Ja
fe3_p	Eisen 3 ⁺	Nein	Ja
for_p	Formiat	Ja	Nein
glc__D_p	D-Glucose	Nein	Ja
h_p	Proton	Ja	Ja
h2o_p	Wasser	Ja	Ja
k_p	Kalium	Nein	Ja
lac__D_p	D-Lactose	Ja	Nein
meoh_p	Methanol	Ja	Nein
mg2_p	Magnesium 2 ⁺	Nein	Ja
mn2_p	Mangan 2 ⁺	Nein	Ja
mobd_p	Molybdän	Nein	Ja
mththf_c	(2R,4S)-2-Methyl-2,3,3,4-Tetrahydroxytetrahydrofuran	Ja	Nein
na1_p	Natrium (nur in <i>iML1515</i>)	Nein	Ja
ni2_p	Nickel 2 ⁺	Nein	Ja
nh4_p	Ammonium	Nein	Ja
o2_p	Sauerstoff	Nein	Ja
oxam_c	Oxamat	Ja	Nein
pi_p	Phosphat	Nein	Ja
sel_p	Selenat (nur in <i>iML1515</i>)	Nein	Ja

(Fortsetzung von Tab. A4)

slnt_p	Selenit (nur in <i>iML1515</i>)	Nein	Ja
so4_p	Sulfat	Nein	Ja
succ_p	Succinat	Ja	Nein
tungs_p	Wolfram (nur in <i>iML1515</i>)	Nein	Ja
zn2_p	Zink 2 ⁺	Nein	Ja

8 Literaturverzeichnis

- [1] L. Michaelis, M.L. Menten (1913): Die Kinetik der Invertinwirkung. *Biochem. Z.* Nr. 49, S. 333–369.
- [2] J.W. Gibbs (1876): On the Equilibrium of Heterogeneous Substances. *Trans. Conn. Acad. Arts Sci.* Nr. 3, S. 108–524.
- [3] A.M. Maxam, W. Gilbert (1977): A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U. S. A.* Vol. 74, Nr. 2, S. 560–564.
- [4] F. Sanger, S. Nicklen, A.R. Coulson (1977): DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* Vol. 74, Nr. 12, S. 5463–5467.
- [5] I. Thiele, B.Ø. Palsson (2010): A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* Vol. 5, Nr. 1, S. 93–121.
- [6] C. Gu u. a. (2019): Current status and applications of genome-scale metabolic models. *Genome Biol.* Vol. 20, Nr. 1, S. 121.
- [7] J.M. Savinell, B.O. Palsson (1992): Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *J. Theor. Biol.* Vol. 154, Nr. 4, S. 421–454.
- [8] M.R. Watson (1984): Metabolic maps for the Apple II. *Biochem. Soc. Trans.* Vol. 12, Nr. 6, S. 1093–1094.
- [9] R. Mahadevan, C.H. Schilling (2003): The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* Vol. 5, Nr. 4, S. 264–276.
- [10] A. Ebrahim, J.A. Lerman, B.O. Palsson, D.R. Hyduke (2013): COBRApy: CONstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* Vol. 7, Nr. 1, S. 74.
- [11] S. Klamt, J. Saez-Rodriguez, E.D. Gilles (2007): Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol.* Vol. 1, Nr. 1, S. 2.
- [12] R. Adadi u. a. (2012): Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters. *PLOS Comput. Biol.* Vol. 8, Nr. 7, S. e1002575.
- [13] B.J. Sánchez u. a. (2017): Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* Vol. 13, Nr. 8, S. 935.
- [14] K.C. Soh, V. Hatzimanikatis (2014): Constraining the Flux Space Using Thermodynamics and Integration of Metabolomics Data. In: Krömer, J.O., Nielsen, L.K., und Blank, L.M. (Hrsg.) *Metabolic Flux Analysis: Methods and Protocols*. Springer, New York, NY. S. 49–63.
- [15] O. Hädicke, A. von Kamp, T. Aydogan, S. Klamt (2018): OptMDFpathway: identification of metabolic pathways with maximal thermodynamic driving force and its application for analyzing the endogenous CO₂ fixation potential of *Escherichia coli*. *PLOS Comput. Biol.* Vol. 14, Nr. 9, S. e1006492.
- [16] N. Jagmann, B. Philipp (2014): Design of synthetic microbial communities for biotechnological production processes. *J. Biotechnol.* Vol. 184, S. 209–218.
- [17] K. Jawed, S.S. Yazdani, M.A.G. Koffas (2019): Advances in the development and application of microbial consortia for metabolic engineering. *Metab. Eng. Commun.* S. e00095.
- [18] N.I. Johns, T. Blazejewski, A.L. Gomes, H.H. Wang (2016): Principles for designing synthetic microbial communities. *Curr. Opin. Microbiol.* Vol. 31, S. 146–153.
- [19] W. Sabra, D. Dietz, D. Tjahjajari, A.-P. Zeng (2010): Biosystems analysis and engineering of microbial consortia for industrial biotechnology. *Eng. Life Sci.* Vol. 10, Nr. 5, S. 407–421.
- [20] J. Shong, M.R. Jimenez Diaz, C.H. Collins (2012): Towards synthetic microbial consortia for bioprocessing. *Curr. Opin. Biotechnol.* Vol. 23, Nr. 5, S. 798–802.
- [21] S.H.J. Chan, M.N. Simons, C.D. Maranas (2017): SteadyCom: Predicting microbial abundances while ensuring community stability. *PLOS Comput. Biol.* Vol. 13, Nr. 5, S. e1005539.
- [22] S. Koch u. a. (2019): RedCom: A strategy for reduced metabolic modeling of complex microbial communities and its application for analyzing experimental datasets from anaerobic digestion. *PLOS Comput. Biol.* Vol. 15, Nr. 2, S. e1006759.
- [23] K.R. Patil, M. Åkesson, J. Nielsen (2004): Use of genome-scale microbial models for metabolic engineering. *Curr. Opin. Biotechnol.* Vol. 15, Nr. 1, S. 64–69.
- [24] A.P. Burgard, P. Pharkya, C.D. Maranas (2003): Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* Vol. 84, Nr. 6, S. 647–657.
- [25] N. Tepper, T. Shlomi (2010): Predicting metabolic engineering knockout strategies for chemical

- production: accounting for competing pathways. *Bioinformatics*. Vol. 26, Nr. 4, S. 536–543.
- [26] K. Jensen u. a. (2019): OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs. *Metab. Eng. Commun.* Vol. 8, S. e00087.
- [27] P. Schneider, A. von Kamp, S. Klamt (2020): An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets. *PLOS Comput. Biol.* Vol. 16, Nr. 7, S. e1008110.
- [28] B.-J. Harder, K. Bettenbrock, S. Klamt (2016): Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*. *Metab. Eng.* Vol. 38, S. 29–37.
- [29] S. Klamt, R. Mahadevan (2015): On the feasibility of growth-coupled product synthesis in microbial strains. *Metab. Eng.* Vol. 30, S. 166–178.
- [30] T. Großkopf u. a. (2016): Metabolic modelling in a dynamic evolutionary framework predicts adaptive diversification of bacteria in a long-term evolution experiment. *BMC Evol. Biol.* Vol. 16, S. 1–12.
- [31] W.R. Harcombe u. a. (2013): The Ability of Flux Balance Analysis to Predict Evolution of Central Metabolism Scales with the Initial Distance to the Optimum. *PLoS Comput. Biol.* Vol. 9, Nr. 6, S. e1003091.
- [32] C. Pál, B. Papp, M.J. Lercher (2005): Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* Vol. 37, Nr. 12, S. 1372–1375.
- [33] D. Machado, S. Andrejev, M. Tramontano, K.R. Patil (2018): Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res.* Vol. 46, Nr. 15, S. 7542–7553.
- [34] O. Dias, M. Rocha, E.C. Ferreira, I. Rocha (2015): Reconstructing genome-scale metabolic models with merlin. *Nucleic Acids Res.* Vol. 43, Nr. 8, S. 3899–3910.
- [35] S.M.D. Seaver u. a. (2021): The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. *Nucleic Acids Res.* Vol. 49, Nr. D1, S. D575–D588.
- [36] E. Noor u. a. (2013): A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. *FEBS Lett.* Vol. 587, Nr. 17, S. 2772–2777.
- [37] E. Klipp, W. Liebermeister, C. Wierling, A. Kowald (2016): *Systems Biology: A Textbook*. John Wiley & Sons.
- [38] S. Klamt, O. Hädicke, A. von Kamp (2014): Stoichiometric and Constraint-Based Analysis of Biochemical Reaction Networks. In: Benner, P. u. a. (Hrsg.) *Large-Scale Networks in Engineering and Life Sciences*. Springer International Publishing. Cham. S. 263–316.
- [39] R. Heinrich, S. Schuster (1996): *Metabolic Control Analysis*. In: Heinrich, R. und Schuster, S. (Hrsg.) *The Regulation of Cellular Systems*. Springer US. Boston, MA. S. 138–291.
- [40] H. Link, D. Christodoulou, U. Sauer (2014): Advancing metabolic models with kinetic information. *Curr. Opin. Biotechnol.* Vol. 29, S. 8–14.
- [41] A. Varma, B.O. Palsson (1994): Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. *Bio/Technology*. Vol. 12, Nr. 10, S. 994–998.
- [42] J.D. Orth, I. Thiele, B.Ø. Palsson (2010): What is flux balance analysis? *Nat. Biotechnol.* Vol. 28, Nr. 3, S. 245–248.
- [43] S. Klamt, S. Müller, G. Regensburger, J. Zanghellini (2018): A mathematical framework for yield (vs. rate) optimization in constraint-based modeling and applications in metabolic engineering. *Metab. Eng.* Vol. 47, S. 153–169.
- [44] M.J. van Hoek, R.M. Merks (2012): Redox balance is key to explaining full vs. partial switching to low-yield metabolism. *BMC Syst. Biol.* Vol. 6, Nr. 1, S. 22.
- [45] M.E. Beber u. a. (2022): eEquilibrator 3.0: a database solution for thermodynamic constant estimation. *Nucleic Acids Res.* Vol. 50, Nr. D1, S. D603–D609.
- [46] E. Noor, H.S. Haraldsdóttir, R. Milo, R.M.T. Fleming (2013): Consistent Estimation of Gibbs Energy Using Component Contributions. *PLOS Comput. Biol.* Vol. 9, Nr. 7, S. e1003098.
- [47] E. Noor u. a. (2014): Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLOS Comput. Biol.* Vol. 10, Nr. 2, S. e1003483.
- [48] O. Hädicke, S. Klamt (2011): Computing complex metabolic intervention strategies using constrained minimal cut sets. *Metab. Eng.* Vol. 13, Nr. 2, S. 204–213.
- [49] A. von Kamp, S. Klamt (2014): Enumeration of Smallest Intervention Strategies in Genome-Scale Metabolic Networks. *PLOS Comput. Biol.* Vol. 10, Nr. 1, S. e1003378.

- [50] K. Ballerstein, A. von Kamp, S. Klamt, U.-U. Haus (2012): Minimal cut sets in a metabolic network are elementary modes in a dual network. *Bioinformatics*. Vol. 28, Nr. 3, S. 381–387.
- [51] B. García-Jiménez, J. Torres-Bacete, J. Nogales (2021): Metabolic modelling approaches for describing and engineering microbial communities. *Comput. Struct. Biotechnol. J.* Vol. 19, S. 226–246.
- [52] H.-S. Song, W.R. Cannon, A.S. Beliaev, A. Konopka (2014): Mathematical modeling of microbial community dynamics: a methodological review. *Processes*. Vol. 2, Nr. 4, S. 711–752.
- [53] A. Succurro, O. Ebenhöf (2018): Review and perspective on mathematical modeling of microbial ecosystems. *Biochem. Soc. Trans.* Vol. 46, Nr. 2, S. 403–412.
- [54] E. Anane, D.C. López C, P. Neubauer, M.N. Cruz Bournazou (2017): Modelling overflow metabolism in *Escherichia coli* by acetate cycling. *Biochem. Eng. J.* Vol. 125, S. 23–30.
- [55] M. Mauri, J.-L. Gouzé, H. de Jong, E. Cinquemani (2020): Enhanced production of heterologous proteins by a synthetic microbial community: Conditions and trade-offs. *PLOS Comput. Biol.* Vol. 16, Nr. 4, S. e1007795.
- [56] S. Benito-Vaquerizo u. a. (2020): Modeling a co-culture of *Clostridium autoethanogenum* and *Clostridium kluyveri* to increase syngas conversion to medium-chain fatty-acids. *bioRxiv*. S. 2020.06.23.167189.
- [57] S.R. Chaganti, D.-H. Kim, J.A. Lalman (2011): Flux balance analysis of mixed anaerobic microbial communities: Effects of linoleic acid (LA) and pH on biohydrogen production. *Int. J. Hydrog. Energy*. Vol. 36, Nr. 21, S. 14141–14152.
- [58] P.S. Losoi, V.P. Santala, S.M. Santala (2019): Enhanced population control in a synthetic bacterial consortium by interconnected carbon cross-feeding. *ACS Synth. Biol.* Vol. 8, Nr. 12, S. 2642–2650.
- [59] N.D. Schwalm u. a. (2019): Developing a microbial consortium for enhanced metabolite production from simulated food waste. *Fermentation*. Vol. 5, Nr. 4, S. 98.
- [60] S. Stolyar u. a. (2007): Metabolic modeling of a mutualistic microbial community. *Mol. Syst. Biol.* Vol. 3, Nr. 1, S. 92.
- [61] S. Koch u. a. (2016): Predicting compositions of microbial communities from stoichiometric models with applications for the biogas process. *Biotechnol. Biofuels*. Vol. 9, Nr. 1, S. 17.
- [62] R.A. Khandelwal u. a. (2013): Community flux balance analysis for microbial consortia at balanced growth. *PLOS ONE*. Vol. 8, Nr. 5, S. e64567.
- [63] T. Escherich (1885): Die Darmbakterien des Neugeborenen und Säuglings. *Fortschr. Med.* Vol. 3, S. 515–522.
- [64] F.R. Blattner u. a. (1997): The Complete Genome Sequence of *Escherichia coli* K-12. *Science*. Vol. 277, Nr. 5331, S. 1453–1462.
- [65] G. Gottschalk (1986): *Bacterial Metabolism*. Springer. New York, NY.
- [66] M. Basan u. a. (2015): Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*. Vol. 528, Nr. 7580, S. 99–104.
- [67] D.P. Clark (1989): The fermentation pathways of *Escherichia coli*. *FEMS Microbiol. Lett.* Vol. 63, Nr. 3, S. 223–234.
- [68] J. Monod (1949): The Growth of Bacterial Cultures. *Annu. Rev. Microbiol.* Vol. 3, S. 371–394.
- [69] J.S. Edwards, B.O. Palsson (2000): The *Escherichia coli* MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci.* Vol. 97, Nr. 10, S. 5528–5533.
- [70] J.D. Orth u. a. (2011): A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism —2011. *Mol. Syst. Biol.* Vol. 7, Nr. 1, S. 535.
- [71] O. Hädicke, S. Klamt (2017): EColiCore2: a reference network model of the central metabolism of *Escherichia coli* and relationships to its genome-scale parent model. *Sci. Rep.* Vol. 7, S. 39647.
- [72] P. Erdrich, R. Steuer, S. Klamt (2015): An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. *BMC Syst. Biol.* Vol. 9, Nr. 1, S. 48.
- [73] J.M. Monk u. a. (2017): iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* Vol. 35, Nr. 10, S. 904–908.
- [74] M. Kanehisa, S. Goto (2000): KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* Vol. 28, Nr. 1, S. 27–30.
- [75] R. Caspi u. a. (2010): The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* Vol. 38, Nr. Database issue, S. D473–D479.
- [76] M. Ganter u. a. (2013): MetaNetX.org: a website and repository for accessing, analysing and

- manipulating metabolic networks. *Bioinformatics*. Vol. 29, Nr. 6, S. 815–816.
- [77] Z.A. King u. a. (2016): BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* Vol. 44, Nr. D1, S. D515–D522.
- [78] M. Hucka u. a. (2003): The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*. Vol. 19, Nr. 4, S. 524–531.
- [79] B.J. Bornstein, S.M. Keating, A. Jouraku, M. Hucka (2008): LibSBML: an API library for SBML. *Bioinforma. Oxf. Engl.* Vol. 24, Nr. 6, S. 880–881.
- [80] S. Hoops u. a. (2006): COPASI—a COMplex PATHway SIMulator. *Bioinformatics*. Vol. 22, Nr. 24, S. 3067–3074.
- [81] K. Choi u. a. (2018): Tellurium: An extensible python-based modeling environment for systems and synthetic biology. *Biosystems*. Vol. 171, S. 74–79.
- [82] S.A. Becker u. a. (2007): Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* Vol. 2, Nr. 3, S. 727–738.
- [83] I. Domenzain u. a. (2022): Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nat. Commun.* Vol. 13, Nr. 1, S. 3766.
- [84] J.G.R. Cardoso u. a. (2018): Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories. *ACS Synth. Biol.* Vol. 7, Nr. 4, S. 1163–1166.
- [85] S. Thiele u. a. (2022): CNAPy: a CellNetAnalyzer GUI in Python for analyzing and designing metabolic networks. *Bioinformatics*. Vol. 38, Nr. 5, S. 1467–1469.
- [86] P. Schneider, P.S. Bekiaris, A. von Kamp, S. Klamt (2022): StrainDesign: a comprehensive Python package for computational design of metabolic networks. *Bioinforma. Oxf. Engl.* Vol. 38, Nr. 21, S. 4981–4983.
- [87] J.C. Muriel, C. Long, N. Sonnenschein (2023): Geckopy 3.0: enzyme constraints, thermodynamics constraints and omics integration in python, <https://www.biorxiv.org/content/10.1101/2023.03.20.533446v1>.
- [88] L. Jeske u. a. (2019): BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* Vol. 47, Nr. D1, S. D542–D549.
- [89] U. Wittig u. a. (2012): SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Res.* Vol. 40, Nr. D1, S. D790–D796.
- [90] S. Federhen (2012): The NCBI Taxonomy database. *Nucleic Acids Res.* Vol. 40, Nr. D1, S. D136–D143.
- [91] UniProt Consortium (2019): UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* Vol. 47, Nr. D1, S. D506–D515.
- [92] K. Valgepea, K. Adamberg, A. Seiman, R. Vilu (2013): Escherichia coli achieves faster growth by increasing catalytic and translation rates of proteins. *Mol. Biosyst.* Vol. 9, Nr. 9, S. 2344–2358.
- [93] J.M. Monk u. a. (2016): Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes. *Cell Syst.* Vol. 3, Nr. 3, S. 238–251.e12.
- [94] S. Boecker u. a. (2019): Broadening the Scope of Enforced ATP Wasting as a Tool for Metabolic Engineering in *Escherichia coli*. *Biotechnol. J.* S. 1800438.
- [95] A. Goelzer, V. Fromion, G. Scorletti (2011): Cell design in bacteria as a convex optimization problem. *Automatica*. Vol. 47, Nr. 6, S. 1210–1218.
- [96] J.A. Lerman u. a. (2012): In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* Vol. 3, Nr. 1, S. 929.
- [97] A. Bulović u. a. (2019): Automated generation of bacterial resource allocation models. *Metab. Eng.* Vol. 55, S. 12–22.
- [98] C.J. Lloyd u. a. (2018): COBRAME: A computational framework for genome-scale models of metabolism and gene expression. *PLoS Comput. Biol.* Vol. 14, Nr. 7, S. e1006302.
- [99] G.W. Luli, W.R. Strohl (1990): Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. *Appl. Environ. Microbiol.* Vol. 56, Nr. 4, S. 1004–1011.
- [100] D.H. de Groot u. a. (2020): The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cell. Mol. Life Sci.* Vol. 77, Nr. 3, S. 441–453.
- [101] J.C. Xavier u. a. (2021): The metabolic network of the last bacterial common ancestor. *Commun. Biol.* Vol. 4, Nr. 1, S. 1–10.
- [102] A. Flamholz, E. Noor, A. Bar-Even, R. Milo (2012): eQuilibrator—the biochemical thermodynamics

- calculator. *Nucleic Acids Res.* Vol. 40, Nr. D1, S. D770–D775.
- [103] B.D. Bennett u. a. (2009): Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* Vol. 5, Nr. 8, S. 593–599.
- [104] J.E. Goldford, A.B. George, A.I. Flamholz, D. Segrè (2022): Protein cost minimization promotes the emergence of coenzyme redundancy. *Proc. Natl. Acad. Sci.* Vol. 119, Nr. 14, S. e2110787119.
- [105] Z.A. King, A.M. Feist (2013): Optimizing Cofactor Specificity of Oxidoreductase Enzymes for the Generation of Microbial Production Strains—OptSwap. *Ind. Biotechnol.* Vol. 9, Nr. 4, S. 236–246.
- [106] P. Salvy u. a. (2019): pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis. *Bioinformatics.* Vol. 35, Nr. 1, S. 167–169.
- [107] S. Mitchell, M. O’Sullivan, I. Dunning (2011): PuLP: a linear programming toolkit for python. *Univ. Auckl. Auckl. N. Z.* Vol. 65,.
- [108] G. Zhu, G.B. Golding, A.M. Dean (2005): The Selective Cause of an Ancient Adaptation. *Science.* Vol. 307, Nr. 5713, S. 1279–1282.
- [109] F.P. Bologna, C.S. Andreo, M.F. Drincovich (2007): *Escherichia coli* Malic Enzymes: Two Isoforms with Substantial Differences in Kinetic Properties, Metabolic Regulation, and Structure. *J. Bacteriol.* Vol. 189, Nr. 16, S. 5937–5946.
- [110] J. Zhang u. a. (2015): Determination of the Cytosolic NADPH/NADP Ratio in *Saccharomyces cerevisiae* using Shikimate Dehydrogenase as Sensor Reaction. *Sci. Rep.* Vol. 5, Nr. 1, S. 12846.
- [111] T. Obata (2020): Toward an evaluation of metabolite channeling in vivo. *Curr. Opin. Biotechnol.* Vol. 64, S. 55–61.
- [112] M. Pavan u. a. (2022): Advances in systems metabolic engineering of autotrophic carbon oxide-fixing biocatalysts towards a circular economy. *Metab. Eng.* Vol. 71, S. 117–141.
- [113] E. Armingol, E. Tobar, R. Cabrera (2018): Understanding the impact of the cofactor swapping of isocitrate dehydrogenase over the growth phenotype of *Escherichia coli* on acetate by using constraint-based modeling. *PLOS ONE.* Vol. 13, Nr. 4, S. e0196182.
- [114] B.K.-S. Chung u. a. (2013): Genome-scale in silico modeling and analysis for designing synthetic terpenoid-producing microbial cell factories. *Chem. Eng. Sci.* Vol. 103, S. 100–108.
- [115] A. Ghosh, H. Zhao, N.D. Price (2011): Genome-Scale Consequences of Cofactor Balancing in Engineered Pentose Utilization Pathways in *Saccharomyces cerevisiae*. *PLOS ONE.* Vol. 6, Nr. 11, S. e27316.
- [116] Z.A. King, A.M. Feist (2014): Optimal cofactor swapping can increase the theoretical yield for chemical production in *Escherichia coli* and *Saccharomyces cerevisiae*. *Metab. Eng.* Vol. 24, S. 117–128.
- [117] M. Lakshmanan, K. Yu, L. Koduru, D.-Y. Lee (2015): In silico model-driven cofactor engineering strategies for improving the overall NADP(H) turnover in microbial cell factories. *J. Ind. Microbiol. Biotechnol.* Vol. 42, Nr. 10, S. 1401–1414.
- [118] J.K.B. Cahn u. a. (2017): A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases. *ACS Synth. Biol.* Vol. 6, Nr. 2, S. 326–333.
- [119] S. Bastian u. a. (2011): Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metab. Eng.* Vol. 13, Nr. 3, S. 345–352.
- [120] M. Bouzon u. a. (2021): Change in Cofactor Specificity of Oxidoreductases by Adaptive Evolution of an *Escherichia coli* NADPH-Auxotrophic Strain. *mBio.* Vol. 12, Nr. 4, S. e00329-21.
- [121] J.M. Smith, E. Szathmary (1997): The major transitions in evolution. OUP Oxford.
- [122] S. Pande u. a. (2014): Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. *ISME J.* Vol. 8, Nr. 5, S. 953–962.
- [123] H. Zhang, X. Wang (2016): Modular co-culture engineering, a new approach for metabolic engineering. *Metab. Eng.* Vol. 37, S. 114–121.
- [124] C.R. Lee u. a. (2018): Co-culture-based biological carbon monoxide conversion by *Citrobacter amalonaticus* Y19 and *Sporomusa ovata* via a reducing-equivalent transfer mediator. *Bioresour. Technol.* Vol. 259, S. 128–135.
- [125] K. Zhou, K. Qiao, S. Edgar, G. Stephanopoulos (2015): Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* Vol. 33, Nr. 4, S. 377–383.
- [126] H.-D. Shin, S. McClendon, T. Vo, R.R. Chen (2010): *Escherichia coli* binary culture engineered for direct fermentation of hemicellulose to a biofuel. *Appl. Environ. Microbiol.* Vol. 76, Nr. 24, S. 8150–

- [127] J.A. Jones u. a. (2016): Experimental and computational optimization of an Escherichia coli co-culture for the efficient production of flavonoids. *Metab. Eng.* Vol. 35, S. 55–63.
- [128] J.A. Jones u. a. (2017): Complete biosynthesis of anthocyanins using E. coli polycultures. *mBio*. Vol. 8, Nr. 3,.
- [129] E. Harvey, J. Heys, T. Gedeon (2014): Quantifying the effects of the division of labor in metabolic pathways. *J. Theor. Biol.* Vol. 360, S. 222–242.
- [130] J.-U. Kreft, B.M. Griffin, R. González-Cabaleiro (2020): Evolutionary causes and consequences of metabolic division of labour: why anaerobes do and aerobes don't. *Curr. Opin. Biotechnol.* Vol. 62, S. 80–87.
- [131] S.R. Lindemann u. a. (2016): Engineering microbial consortia for controllable outputs. *ISME J.* Vol. 10, Nr. 9, S. 2077–2084.
- [132] G.W. Roell u. a. (2019): Engineering microbial consortia by division of labor. *Microb. Cell Factories*. Vol. 18, Nr. 1, S. 35.
- [133] R. Tsoi u. a. (2018): Metabolic division of labor in microbial systems. *Proc. Natl. Acad. Sci.* Vol. 115, Nr. 10, S. 2526–2531.
- [134] G. Wu u. a. (2016): Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol.* Vol. 34, Nr. 8, S. 652–664.
- [135] N.E. Lewis u. a. (2010): Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* Vol. 6, Nr. 1, S. 390.
- [136] C. Diener, S.M. Gibbons, O. Resendis-Antonio (2020): MICOM: Metagenome-scale modeling to infer metabolic interactions in the gut microbiota. *mSystems*. Vol. 5, Nr. 1,.
- [137] J.J. Hamilton, V. Dwivedi, J.L. Reed (2013): Quantitative assessment of thermodynamic constraints on the solution space of genome-scale metabolic models. *Biophys. J.* Vol. 105, Nr. 2, S. 512–522.
- [138] B. García-Jiménez, J.L. García, J. Nogales (2018): FLYCOP: metabolic modeling-based analysis and engineering microbial communities. *Bioinformatics*. Vol. 34, Nr. 17, S. i954–i963.
- [139] M. Thommes u. a. (2019): Designing Metabolic Division of Labor in Microbial Communities. *mSystems*. Vol. 4, Nr. 2, S. e00263-18.
- [140] N. Venayak, A. von Kamp, S. Klamt, R. Mahadevan (2018): MoVE identifies metabolic valves to switch between phenotypic states. *Nat. Commun.* Vol. 9, Nr. 1, S. 5332.
- [141] C.J. Hartline, A.C. Schmitz, Y. Han, F. Zhang (2021): Dynamic control in metabolic engineering: Theories, tools, and applications. *Metab. Eng.* Vol. 63, S. 126–140.
- [142] L. Wang, V. Upadhyay, C.D. Maranas (2021): dGPredictor: Automated fragmentation method for metabolic reaction free energy prediction and de novo pathway design. *PLOS Comput. Biol.* Vol. 17, Nr. 9, S. e1009448.
- [143] D. Heckmann u. a. (2018): Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat. Commun.* Vol. 9, Nr. 1, S. 5252.
- [144] X. Yang u. a. (2021): Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models. *Metab. Eng.* Vol. 67, S. 133–144.

9 Liste der referenzierten elektronischen Repositorys

Im Rahmen dieser Arbeit wurden folgende Internet-Repositorys wie folgt referenziert:

- *Kapitel 3*: AutoPACMEN mit der Referenzimplementierung des Algorithmus sowie Erstellungs- und Testskripten bezüglich iJO1366*
<https://github.com/klamt-lab/autopacmen>
- *Kapitel 4*: TCOSA, mit der Referenzimplementierung des Algorithmus und <https://www.github.com/klamt-lab/TCOSA>
Ergänzender Datensatz 1
<https://github.com/Paulocracy/Dissertationsdaten> oder gespiegelt unter
<https://doi.org/10.5281/zenodo.10556528>
- *Kapitel 5*: Das kinetische Beispielmodell
<https://github.com/Paulocracy/kinetic-modelling> oder gespiegelt unter
<https://doi.org/10.5281/zenodo.10556533>
CommModelPy, zusammen mit der Erstellung der *E. coli*-Gemeinschaftsmodelle
<https://github.com/klamt-lab/CommModelPy>
ASTHERISC Package, mit der Referenzimplementierung des ASTHERISC-Algorithmus
und der Anwendung auf die *E. coli*-Gemeinschaftsmodelle
<https://github.com/klamt-lab/astheriscPackage>

10 Liste der Veröffentlichungen in Bezug auf diese Arbeit

Ein Großteil der Resultate wurde in den folgenden Publikationen wie genannt veröffentlicht:

- Kapitel 3: Bekiaris & Klamt, 2020. Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*, 21(1), 1-13.
- Kapitel 5, außer das nicht in dieser Publikation berücksichtigte Unterkapitel 5.2 und alle Referenzen darauf: Bekiaris & Klamt, 2021. Designing microbial communities to maximize the thermodynamic driving force for the production of chemicals. *PloS Computational Biology*, 17(6), e1009093.
- Kapitel 4: Bekiaris & Klamt, 2023. Network-wide thermodynamic constraints shape NAD(P)H cofactor specificity of biochemical reactions. *Nature Communications* 14, 4660.

Weitere Publikationen unter Mitwirkung des Autors, die in einem entfernteren Zusammenhang zur Dissertation stehen, sind:

- Thiele, von Kamp, Bekiaris, Schneider, & Klamt, 2022. CNApy: a *CellNetAnalyzer* GUI in Python for analyzing and designing metabolic networks. *Bioinformatics*, 38(5), 1467-1469.
- Schneider, Bekiaris, von Kamp & Klamt, 2022. StrainDesign: a comprehensive Python package for computational design of metabolic networks. *Bioinformatics*, 38(21), 4981-4983.

Die in diesen beiden Publikationen verwendeten Programmpakete wurden u.a. zum Testen und zur Visualisierung von Aspekten der Modelle, die in dieser Dissertation vorkommen, verwendet.