

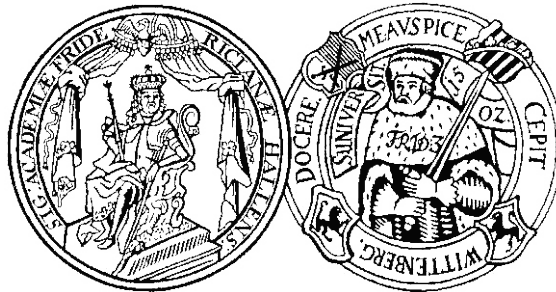
# Integrative Analyse der Struktur und Dynamik metabolischer Modelle

Dissertation  
Zur Erlangung des  
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät III  
Agrar- und Ernährungswissenschaften,  
Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg



vorgelegt von  
Frau Hartmann, Anja  
Geb. am 04.06.1983 in Quedlinburg

Gutachter: Prof. Dr. Falk Schreiber  
Prof. Dr. Ina Koch

Verteidigung am: 20.04.2015

# Danksagung

Mein besonderer Dank gilt meinem Doktorvater Prof. Dr. Falk Schreiber, der mit Rat und Zuspruch dazu beigetragen hat, dass mein wissenschaftliches Arbeit in Veröffentlichungen und dieser Doktorarbeit münden konnten.

Ebenso gebührt mein Dank der ehemaligen Arbeitsgruppe Pflanzenbioinformatik für moralische, fachliche und technische Unterstützung und die produktive Zusammenarbeit in den letzten Jahren. Ein besonderer Dank gebührt Tobias Czauderna für seine Hilfsbereitschaft, vielseitigen Denkanstöße und die Einführung in SBGN. Dr. Eva Grafahrend-Belau danke ich für die geduldige Einführung in die Flussbilanzanalyse.

Bedanken möchte ich mich für die Motivation und fachlichen, sowie nicht-fachlichen Diskussionen bei Dr. Uwe Scholz, Dr. Astrid Junker, Prof. Dr. Björn H. Junker und Dr. Hendrik Rohn. Für das Korrekturlesen möchte ich mich bei Prof. Dr. Björn H. Junker, Dr. Matthias Lange, Cornelia Aschemann und Thomas Thimm bedanken.

Vor allem möchte ich mich bei meinem Freund Kevin Pucknat für seine enorme Geduld, Unterstützung und Verständnis während der Entstehung dieser Arbeit bedanken. Ohne die liebevolle Unterstützung meiner Familie wäre diese Arbeit nicht entstanden. Sie haben mir eine erfolgreiche und sorgenfreie Promotionszeit durch moralische Unterstützung geboten und standen mir bei Schwierigkeiten immer zur Seite. Ich danke euch von Herzen dafür!





# Kurzdarstellung

Die Charakterisierung von Verhaltens- und Funktionsweisen biologischer Systeme stellt aufgrund vielfältiger biochemischer Interaktionen eine große Herausforderung dar. Ansätze zur Modellierung und Analyse werden entwickelt, um die komplexen Mechanismen auf Systemebene zu verstehen. Verschiedene Modellierungsformalismen ermöglichen die Analyse metabolischer Modelle abhängig von der zu lösenden Fragestellung, dem biochemischen Wissen und der Verfügbarkeit experimenteller Daten.

Gegenstand der vorliegenden Arbeit ist die Entwicklung einer Methodik zur integrativen Analyse der Struktur und Dynamik qualitativer und quantitativer metabolischer Modelle. Unter Verwendung verschiedener Modellierungsformalismen wird das metabolische Modell aus mehreren Perspektiven analysiert. Die ermittelten strukturellen und dynamischen Eigenschaften werden im Kontext des metabolischen Modells visualisiert. Interaktionstechniken erlauben Exploration und visuelle Analysen, die zu einem umfassenden Verständnis der Verhaltens- und Funktionsweisen des zugrunde liegenden biologischen Systems führen.

Das *System Biology Metabolic Model Framework (SBM<sup>2</sup>-Framework)* realisiert die entwickelte Methodik und wird zur integrativen Analyse der Kulturpflanzen Kartoffel und Mais angewendet.

## Abstract

The characterization of biological systems with respect to their behavior and functionality based on biochemical interactions is a major challenge. To understand these complex mechanisms at systems level modeling approaches are investigated. Different modeling formalisms allow metabolic models to be analyzed depending on the question to be solved, the biochemical knowledge and the availability of experimental data.

The development of a methodology for an integrative analysis of the structure and dynamics represented by qualitative and quantitative metabolic models is subject of the presented work. Using various modeling formalisms the metabolic model is analyzed from different perspectives. Determined structural and dynamic properties are visualized in the context of the metabolic model. Interaction techniques allow the exploration and visual analysis thereby leading to a broader understanding of the behavior and functionality of the underlying biological system.

The *System Biology Metabolic Model Framework (SBM<sup>2</sup>-Framework)* implements the developed methodology and, as an example, is applied for the integrative analysis of the crop plants potato and corn.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>15</b>
1.1	Motivation . . . . .	16
1.2	Zielsetzung . . . . .	17
1.3	Vorgehensweise . . . . .	17
<b>2</b>	<b>Grundlagen</b>	<b>21</b>
2.1	Grundlagen aus der Biochemie . . . . .	21
2.1.1	Merkmale von Lebewesen . . . . .	21
2.1.2	Biochemische Reaktion . . . . .	22
2.1.3	Stoffwechselwege . . . . .	23
2.1.4	Metabolismus . . . . .	23
2.1.5	Beeinflussung des Metabolismus . . . . .	23
2.2	Grundlagen aus der Mathematik und Informatik . . . . .	24
2.2.1	Mengen und Mengenoperationen . . . . .	25
2.2.2	Relationen und Funktionen . . . . .	25
2.2.3	Logik . . . . .	26
2.2.3.1	Aussagenlogik . . . . .	26
2.2.3.2	Prädikatenlogik . . . . .	26
2.2.4	Graphen . . . . .	27
2.3	Grundlagen aus der Systembiologie . . . . .	29
2.3.1	Metabolische Modellierung . . . . .	29
2.3.1.1	Metabolisches Modell . . . . .	29
2.3.1.2	Rekonstruktion metabolischer Modelle . . . . .	31
2.3.1.3	Erweiterung metabolischer Modelle zur Analyse . . . . .	33
2.3.2	Standardisierte Austauschformate für metabolische Modelle . . . . .	34
2.3.3	Visualisierung und Interaktionstechniken . . . . .	35
2.3.3.1	Visualisierung . . . . .	35
2.3.3.2	Interaktionstechniken . . . . .	36
<b>3</b>	<b>Modellierungsformalismen metabolischer Modelle</b>	<b>39</b>
3.1	Stand der Technik . . . . .	39
3.2	Kinetische Modellierung . . . . .	40
3.2.1	Kinetisches Modell . . . . .	40
3.2.2	Analysen . . . . .	40
3.3	Petri-Netz Modellierung . . . . .	42
3.3.1	Petri-Netz Modell . . . . .	42
3.3.2	Analysen . . . . .	45
3.4	Stöchiometrische Modellierung . . . . .	46
3.4.1	Stöchiometrisches Modell . . . . .	47
3.4.2	Analysen . . . . .	47
3.5	Topologische Modellierung . . . . .	49
3.5.1	Topologisches Modell . . . . .	49

3.5.2	Analysen . . . . .	49
3.6	Vergleich der Modellierungsformalismen . . . . .	50
<b>4</b>	<b>Evaluierung bestehender Ansätze zur metabolischen Modellierung</b>	<b>53</b>
4.1	Evaluierungskriterien . . . . .	53
4.1.1	Kinetische Modellierung . . . . .	54
4.1.2	Petri-Netz Modellierung . . . . .	54
4.1.3	Stöchiometrische Modellierung . . . . .	55
4.1.4	Topologische Modellierung . . . . .	55
4.2	Vergleich spezifischer Modellierungsansätze . . . . .	55
4.2.1	Kinetische Modellierung . . . . .	55
4.2.2	Petri-Netz Modellierung . . . . .	57
4.2.3	Stöchiometrische Modellierung . . . . .	58
4.2.4	Topologische Modellierung . . . . .	59
4.3	Vergleich integrierter Modellierungsansätze . . . . .	61
4.4	Erweiterte Anforderungen . . . . .	64
<b>5</b>	<b>Methodik</b>	<b>65</b>
5.1	Metabolisches Modell . . . . .	65
5.1.1	Abstraktion . . . . .	66
5.1.1.1	Strukturelle Beschreibung . . . . .	66
5.1.1.2	Funktionelle Beschreibung . . . . .	66
5.1.2	Formalisierung . . . . .	67
5.1.3	Visualisierung . . . . .	70
5.2	Ableitung spezieller Modelle zur Analyse . . . . .	73
5.2.1	Kinetisches Modell . . . . .	73
5.2.1.1	Graph . . . . .	73
5.2.1.2	Analysen . . . . .	73
5.2.2	Petri-Netz Modell . . . . .	74
5.2.2.1	Graph . . . . .	74
5.2.2.2	Analysen . . . . .	75
5.2.3	Stöchiometrisches Modell . . . . .	76
5.2.3.1	Graph . . . . .	76
5.2.3.2	Analysen . . . . .	76
5.2.4	Topologische Modelle . . . . .	77
5.2.4.1	Graphen . . . . .	77
5.2.4.2	Analysen . . . . .	80
5.2.5	Zusammenfassung der Analyseergebnisse . . . . .	80
5.3	Integration und Visualisierung der Analyseergebnisse . . . . .	82
5.3.1	Datenzuordnungsfunktionen . . . . .	82
5.3.1.1	Datenzuordnungsfunktion für numerische Werte . . . . .	82
5.3.1.2	Datenzuordnungsfunktion für Graphenelemente . . . . .	82
5.3.2	Visualisierung . . . . .	83
5.3.2.1	Kinetische Modellierung . . . . .	84
5.3.2.2	Petri-Netz Modellierung . . . . .	84
5.3.2.3	Stöchiometrische Modellierung . . . . .	85
5.3.2.4	Topologische Modellierung . . . . .	86
5.4	Zusammenfassung . . . . .	88
<b>6</b>	<b>Realisierung</b>	<b>91</b>

6.1	System Biology Metabolic Model Framework . . . . .	91
6.1.1	Import und Export metabolischer Modelle . . . . .	92
6.1.2	Visualisierung metabolischer Modelle . . . . .	93
6.1.3	Analyse metabolischer Modelle . . . . .	94
6.1.3.1	Kinetische Analyse . . . . .	94
6.1.3.2	Petri-Netz Analyse . . . . .	94
6.1.3.3	Stöchiometrische Analyse . . . . .	95
6.1.3.4	Topologische Analyse . . . . .	96
6.2	Zusammenfassung . . . . .	96
<b>7</b>	<b>Anwendung</b>	<b>97</b>
7.1	Anwendung verschiedener spezifischer Modelle . . . . .	97
7.1.1	Kinetische Analyse . . . . .	98
7.1.2	Petri-Netz Analyse . . . . .	98
7.1.3	Stöchiometrische Analyse . . . . .	98
7.1.4	Topologische Analyse . . . . .	100
7.1.5	Zusammenfassung . . . . .	101
7.2	Integrative Anwendung eines metabolischen Modells der Kartoffel . . . . .	101
7.2.1	Kinetische Analyse . . . . .	101
7.2.2	Petri-Netz Analyse . . . . .	103
7.2.3	Stöchiometrische Analyse . . . . .	106
7.2.4	Topologische Analyse . . . . .	107
7.2.5	Zusammenfassung . . . . .	110
7.3	Integrative Anwendung metabolischer Modelle der Maispflanze . . . . .	111
7.3.1	Maiskorn-Modell . . . . .	112
7.3.2	Maisblatt-Modell . . . . .	113
7.3.3	Topologische Analyse . . . . .	113
7.3.4	Stöchiometrische Analyse . . . . .	117
7.3.5	Petri-Netz Analyse . . . . .	121
7.3.6	Kinetische Analyse . . . . .	123
7.3.7	Zusammenfassung . . . . .	126
<b>8</b>	<b>Diskussion</b>	<b>129</b>
8.1	Zusammenfassung . . . . .	129
8.2	Diskussion der Ergebnisse . . . . .	130
8.3	Ausblick . . . . .	132
	<b>Literaturverzeichnis</b>	<b>135</b>



# Abbildungsverzeichnis

1.1	Modelle und Modellierungsfomalismen. . . . .	15
1.2	Ziel der vorliegenden Arbeit. . . . .	18
1.3	Struktur der vorliegenden Arbeit. . . . .	19
2.1	Struktur des Kapitels Grundlagen. . . . .	21
2.2	Informationsfluss in der Zelle. . . . .	24
2.3	Zusammenhang zwischen Signaltransduktion, Genregulation und Metabolismus. . . . .	24
2.4	Phylogenetischer Baum metabolischer Modelle. . . . .	30
2.5	Beispiel eines kleinen metabolischen Modells in <i>SBGN-PD</i> . . . . .	31
2.6	Zyklus der Systembiologie. . . . .	34
3.1	Struktur des Kapitels Modellierungsformalismen. . . . .	39
3.2	Herleitung der <i>ODEs</i> zur kinetischen Analyse. . . . .	41
3.3	Herleitung der Gleichungen zur Invariantenanalyse. . . . .	45
3.4	Herleitung zur Berechnung der optimalen Lösung mittels <i>FBA</i> . . . . .	48
3.5	Herleitung der Metabolit- und Reaktionsgraphen. . . . .	50
4.1	Struktur des Kapitels Evaluierung. . . . .	53
5.1	Struktur des Kapitels Methodik. . . . .	65
5.2	Datenstruktur des abstrakten metabolischen Modells. . . . .	70
5.3	Visualisierung in <i>SBGN-PD</i> . . . . .	71
5.4	Visualisierung des <i>abstrakten Graphen</i> in <i>SBGN-PD</i> . . . . .	72
5.5	Ableitung reversibler Reaktionen des <i>abstrakten Graphen</i> . . . . .	75
5.6	Ableitung von inhibierten Reaktionen des <i>abstrakten Graphen</i> . . . . .	77
5.7	Ableitung des <i>abstrakten Graphen</i> in den <i>Metabolitgraphen</i> . . . . .	79
5.8	Ableitung des <i>abstrakten Graphen</i> in den <i>Reaktionsgraphen</i> . . . . .	80
5.9	Visualisierung der Analyseergebnisse der kinetischen Analyse. . . . .	84
5.10	Visualisierung der Analyseergebnisse der Invariantenanalyse. . . . .	85
5.11	Visualisierung der Analyseergebnisse der optimierungsbasierten Analyse. . . . .	86
5.12	Visualisierung der Analyseergebnisse der Zentralitätsanalyse. . . . .	87
5.13	Integrierte Visualisierung der Analyseergebnisse. . . . .	87
5.14	Zusammenfassung der Methodik. . . . .	88
6.1	Architektur des <i>SBM<sup>2</sup>-Frameworks</i> . . . . .	92
7.1	Struktur des Kapitels Anwendung. . . . .	97
7.2	Kinetische Analyse des Zuckerrohrhalm-Modells. . . . .	99
7.3	Petri-Netz Analyse des einkeimblättrigen Pflanzen-Modells. . . . .	100
7.4	Stöchiometrische Analyse des Gerstensamen-Modells. . . . .	102
7.5	Kinetische Analyse des Kartoffel-Modells. . . . .	104
7.6	Petri-Netz Analyse des Kartoffel-Modells. . . . .	105

7.7	Stöchiometrische Analyse des Kartoffel-Modells. . . . .	108
7.8	Topologische Analyse des Kartoffel-Modells. . . . .	109
7.9	Integrative Analyse des Kartoffel-Modells. . . . .	110
7.10	Schematische Darstellung des Maiskorn-Modells. . . . .	112
7.11	Querschnitt durch das Blatt einer Maispflanze. . . . .	113
7.12	Schematische Darstellung des Maisblatt-Modells. . . . .	114
7.13	Topologische Analyse der Mais-Modelle. . . . .	115
7.14	Integrative Analyse des Maiskorn-Modells. . . . .	118
7.15	Flussbilanzanalyse des Maiskorn-Modells nach Alonso. . . . .	119
7.16	Integrative Analyse des Maisblatt-Modells. . . . .	120
7.17	Integrative Analyse des Maisblatt-Modells (Fortsetzung 2). . . . .	124
7.18	Integrative Analyse des Maisblatt-Modells (Fortsetzung 3). . . . .	126
8.1	Struktur des Kapitels Diskussion. . . . .	129
8.2	Zusammenfassung. . . . .	132



# Tabellenverzeichnis

3.1	Bewertung der vorgestellten Modellierungsformalismen. . . . .	51
4.1	Vergleich kinetischer Anwendungen. . . . .	56
4.2	Vergleich Petri-Netz Anwendungen. . . . .	57
4.3	Vergleich stöchiometrischer Anwendungen. . . . .	59
4.4	Vergleich topologischer Anwendungen. . . . .	60
4.5	Vergleich integrierter Anwendungen. . . . .	62
5.1	Zusammenfassung typischer Analysemethoden. . . . .	81
5.2	Integration unterschiedlicher Modellierungsformalismen. . . . .	89
6.1	Transformation von biologischem Wissen in <i>SBGN-PD</i> , <i>SBML</i> und <i>PNML</i> . . . . .	93

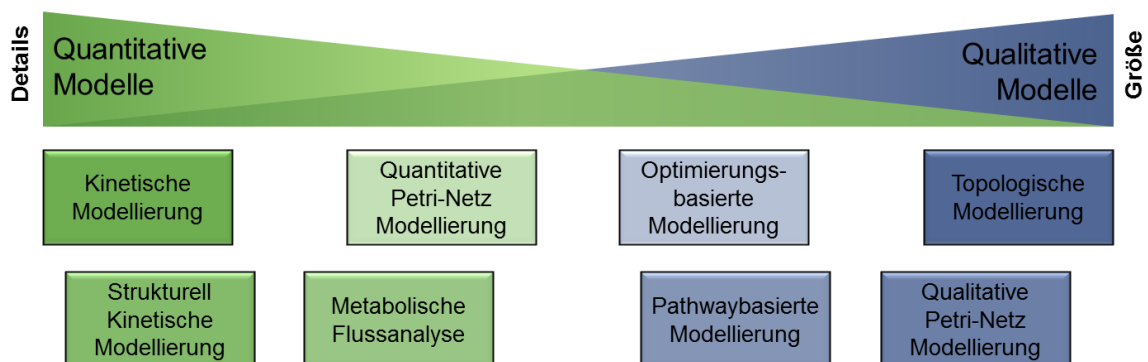


# 1 Einleitung

Der Metabolismus ist die Gesamtheit aller in biologischen Systemen ablaufenden biochemischen Prozesse und bildet somit die Essenz des Lebens. Vielfältige biochemische Interaktionen bilden ein komplexes metabolisches Netzwerk. Zur Charakterisierung von Verhaltens- und Funktionsweisen biologischer Systeme werden deshalb Ansätze zur Modellierung und Analyse angewendet. Die metabolische Modellierung ermöglicht die Entwicklung von Modellen und die Anwendung verschiedener Modellierungsformalismen, um die komplexen Mechanismen auf Systemebene zu verstehen.

Abhängig von der zu lösenden Fragestellung, dem biochemischen Wissen und der Verfügbarkeit experimenteller Daten können biologische Systeme durch verschiedene metabolische Modelle repräsentiert werden. Folglich entstehen große, meist qualitative Modelle und kleine, eher quantitative Modelle [153]. Weiterhin kann in einem iterativen Prozess aus einem qualitativen Modell durch sukzessive Integration detaillierter Experimentaldaten ein quantitatives Modell entstehen.

Unterschiedliche Modellgrößen und enthaltene Details ermöglichen die Analyse struktureller und dynamischer Eigenschaften unter Verwendung verschiedener Modellierungsformalismen. Diese bringen vielfältige Analysemethoden mit sich. Sie bilden die Grundlage metabolische Modelle aus mehreren Perspektiven zu untersuchen und ergänzende Erkenntnisse hervorzubringen. Abbildung 1.1 stellt den Zusammenhang zwischen Modellen und einer Auswahl möglicher Modellierungsformalismen dar.



**Abbildung 1.1:** Biologische Systeme werden durch verschiedene metabolische Modelle repräsentiert. Abhängig von detailliertem Wissen über Interaktionsmechanismen entstehen kleine, eher quantitative Modelle und große, meist qualitative Modelle. Verschiedene Modellierungsformalismen verwenden unterschiedliche Modelle (Zusammenhang farblich gekennzeichnet), und liefern ergänzende Erkenntnisse über das zugrunde liegende biologische System.

Die *kinetische Modellierung* basiert auf einer detaillierten, quantitativen Beschreibung eines nahezu vollständig charakterisierten biologischen Systems. Die *strukturell kinetische Modellierung* erfordert weniger detaillierte Informationen als die kinetische Modellierung und basiert auf unterschiedlichen Experimentaldaten. Unter Verwendung von Markierungsdaten werden intrazelluläre Stoffflüsse durch die *metabolische Flussanalyse* bestimmt. Zur Untersuchung nebenläufiger, asynchroner und verteilter Systeme wird die

*quantitative Petri-Netz Modellierung* verwendet. Diese vier Modellierungsformalismen dienen der Analyse dynamischen Verhaltens und der quantitativen Vorhersage von Eigenschaften des zugrunde liegenden biologischen Systems. Aufgrund der notwendigen und meist kompliziert zu ermittelnden Menge von Experimentaldaten sind diese Formalismen oftmals für kleine Modelle limitiert.

Die *optimierungsbasierte Modellierung* nutzt das Konzept der Optimierung unter Einbeziehung von Nebenbedingungen, um quantitative Vorhersagen basierend auf statischen Beschreibungen des Systems zu treffen. Zur Ermittlung nicht redundanter Stoffwechselwege in metabolischen Modellen werden bei der *pathwaybasierten Modellierung* alle möglichen Reaktionswege analysiert. Verschiedene statische und dynamische Eigenschaften werden unter Verwendung der *qualitativen Petri-Netz Modellierung* charakterisiert. Die *topologische Modellierung* erfasst die Struktur des Modells durch die Analyse von Interaktionsmustern zwischen den Modellkomponenten. Diese vier Modellierungsformalismen können aufgrund der wenigen notwendigen Experimentaldaten auf große Modelle angewendet werden. Das dynamische Verhalten kann aber nur eingeschränkt analysiert werden.

Je detaillierter ein metabolisches Modell beschrieben ist, desto präzisere Modellierungsformalismen können angewendet werden (z. B. kinetische Modellierung). Umgekehrt kann ein detailliert beschriebenes metabolisches Modell mit Modellierungsformalismen untersucht werden, die weniger detailliertes Wissen bedürfen (z. B. topologische Modellierung).

## 1.1 Motivation

Das Ziel der Systembiologie im Hinblick auf den Metabolismus ist die Verhaltens- und Funktionsweisen biologischer Systeme mit deren komplexen Interaktionen auf Systemebene zu verstehen. Zur Aufklärung dieser komplexen Zusammenhänge wurden Methoden entwickelt, die verschiedene Modellierungsformalismen zur Modellanalyse nutzen. Aufgrund der meist unzureichenden Experimentaldaten werden unterschiedliche Ansätze verfolgt. Zum einen werden qualitative Modelle unter Verwendung von berechneten Analyseergebnissen erweitert, um eine darauf folgende quantitative Analyse durchführen zu können und zum anderen werden Modelle in der Größe reduziert (entweder werden mehrere Reaktionen zu einer zusammengefasst oder ein Modell wird in kleinere Modelle aufgeteilt), um weniger Experimentaldaten zur quantitativen Analyse zu verwenden.

In vielen Herangehensweisen [1, 14, 24, 28, 75, 102, 149, 150] wird eine Analysemethode der optimierungsbasierten Modellierung, die *Flussbilanzanalyse* (kurz *FBA*), verwendet. Die *FBA* berechnet den Metabolitfluss (Flussverteilung) durch das metabolische Modell. Diese Flussverteilung wird zur kinetischen Modellierung verwendet. Die *Petri-Netz Modellierung* wird von Chen [23] und Gilbert [55] zur Modellreduktion angewendet, um weniger Experimentaldaten zur kinetischen Modellierung zu verwenden. Eine Kombination beider Ansätze verfolgt Machado [107], indem zunächst das Modell reduziert wird. Anschließend wird unter Verwendung von Analyseergebnissen aus der *FBA*, die kinetische Modellierung durchgeführt. Weiterhin wird ein integrativer Ansatz von Heiner [68] vorgestellt, der ausschließlich den *Petri-Netz Formalismus* nutzt, um qualitative und quantitative Analysen anzuwenden.

Bisher existieren allerdings kaum Herangehensweisen zur integrativen Analyse metabolischer Modelle, die verschiedene etablierte Modellierungsformalismen unterstützen und somit ergänzende Erkenntnisse hervorbringen.

## 1.2 Zielsetzung

Das Ziel dieser Arbeit ist es, das Potential verschiedener etablierter Modellierungsformalismen metabolischer Modelle zu nutzen. Dazu wird eine Methodik entwickelt, die die Integration bisher nicht kombinierter, unterschiedlicher Modellierungsformalismen zur Analyse struktureller und dynamischer Eigenschaften, bereitstellt. Damit ein breites Analysespektrum zur Verfügung steht, werden in der hier vorliegenden Arbeit die kinetische, Petri-Netz (quantitativ und qualitativ), stöchiometrische (strukturell und optimierungsbasiert) und topologische Modellierung verwendet.

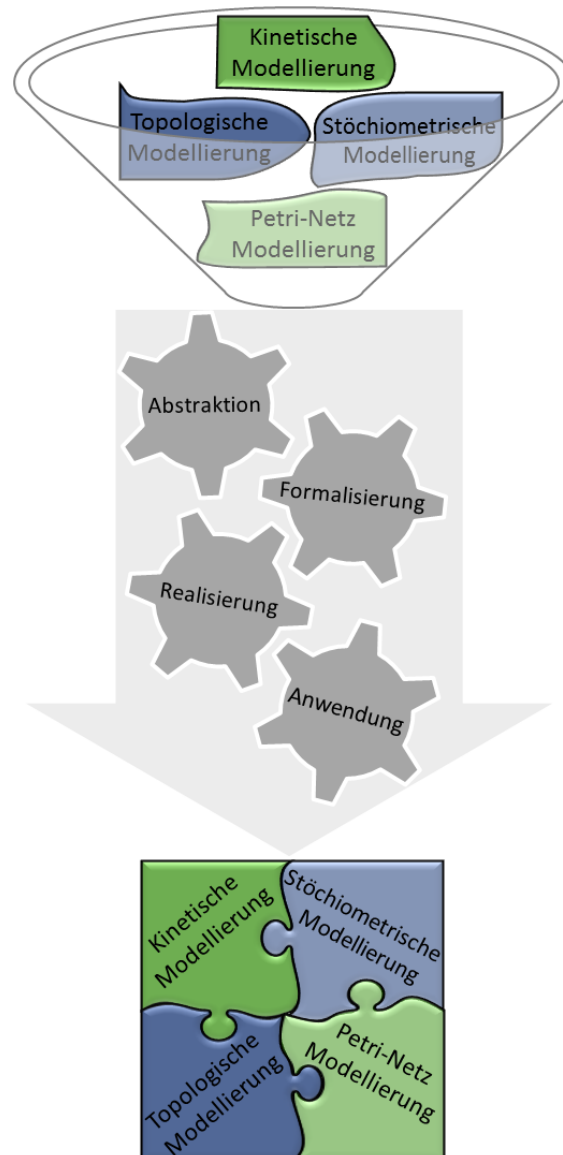
Diese Methodik erlaubt zum einen die Analyse verschiedener qualitativer und quantitativer Modelle mit spezifischen Formalismen, für die diese Modelle rekonstruiert wurden. Zum anderen kann ein metabolisches Modell in einer Anwendung mit verschiedenen Modellierungsformalismen analysiert werden, ohne die graphische Repräsentation anzupassen. Die Integration ermöglicht zudem die Verwendung verschiedener Analyseergebnisse in anderen Modellierungsformalismen. Kann ein metabolisches Modell aufgrund unzureichender Experimentaldaten nicht quantitativ, z. B. kinetisch analysiert werden, wird durch die Anwendung anderer Formalismen eine Analyse durchgeführt, um die fehlenden Daten zu ersetzen. Anstelle fehlender Experimentaldaten für eine oder mehrere Reaktionen im Modell wird somit eine Approximation unter Verwendung anderer Analyseergebnisse ermittelt. Auf diese Weise können trotz fehlender detaillierter Beschreibungen mit dieser Methodik dennoch Rückschlüsse auf die Verhaltens- und Funktionsweise des zugrunde liegenden biologischen Systems gezogen werden.

Weiterhin ist es empfehlenswert, ein bereits detailliert quantitativ beschriebenes metabolisches Modell unter Verwendung zusätzlicher Modellierungsformalismen aus verschiedenen Perspektiven zu analysieren. Die sich ergänzenden Analyseergebnisse und auch die Integration von Analyseergebnissen in anderen Formalismen zur weiterführenden Analyse, unterstützen vor allem die Interpretation der strukturellen und dynamischen Eigenschaften des Modells. Zusätzlich führt die visuelle Integration ermittelter Analyseergebnisse in entsprechende Elemente des metabolischen Modells zu einer Kombination von Modellstruktur und -dynamik. Geeignete Interaktionstechniken unterstützen die Exploration und Interpretation erhobener Ergebnisse und führen zu einem umfassenden Verständnis des zugrunde liegenden biologischen Systems.

Der Fokus dieser Arbeit liegt zusammenfassend in der Entwicklung einer Methodik zur Abstraktion, Formalisierung, Realisierung und Anwendung voneinander unabhängiger, etablierter Modellierungsformalismen, siehe Abbildung 1.2. Es wird damit ein genereller Ansatz zur Integration bisher nicht kombinierbarer Modellierungsformalismen verfolgt. Diese Methodik nutzt das Potential der Formalismen zur integrativen Analyse von Modellen. Infolgedessen werden Datenlücken in Modellen geschlossen, sodass Analyseergebnisse im Kontext von Modellen integrativ exploriert und interpretiert werden. Schließlich werden somit neben strukturellen auch dynamische Eigenschaften des zugrunde liegenden biologischen Systems abgeleitet.

## 1.3 Vorgehensweise

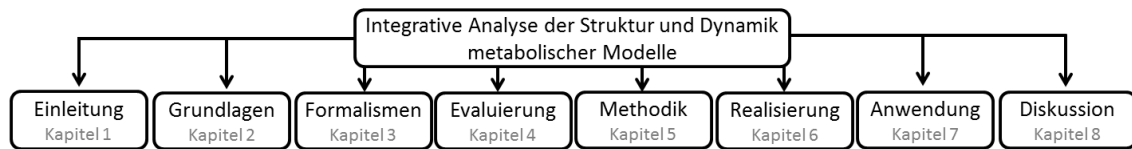
Im Folgenden wird die Vorgehensweise in der vorliegenden Arbeit entlang der Struktur, siehe Abbildung 1.3 beschrieben. Zunächst werden grundlegende Begriffe zur metabolischen Modellierung aus der Biochemie, Mathematik, Informatik und Systembiologie in Kapitel 2 erläutert. Die verwendeten Modellierungsformalismen



**Abbildung 1.2:** Der Fokus dieser Arbeit liegt in der Entwicklung einer Methodik zur Abstraktion, Formalisierung, Realisierung und Anwendung voneinander unabhängiger, etablierter Modellierungsformalismen mit dem Ziel der integrativen Analyse metabolischer Modelle.

werden in Kapitel 3 präsentiert. In Kapitel 4 erfolgt die Evaluierung existierender Computerprogramme, die die Modellierungsformalismen spezifisch als auch integriert unterstützen. Die erweiterten Anforderungen an eine integrierte Anwendung mit einheitlicher Visualisierung werden dargelegt. Die Methodik in Kapitel 5 beinhaltet eine formale Herleitung eines abstrakten metabolischen Modells. Der somit entstehende abstrakte Graph wird zur Verwendung verschiedener Modellierungsformalismen abgeleitet. Es resultieren unterschiedliche Analyseergebnisse, die in eine einheitliche Visualisierung des metabolischen Modells integriert werden. Die Realisierung der Methodik ist in Kapitel 6 in Form der Implementierung *System Biology Metabolic Modell Framework* (*SBM<sup>2</sup>-Framework*) repräsentiert. In Kapitel 7 wird die Anwendung der Methodik anhand von möglichen Vorgehensweisen beschrieben. Zunächst werden spezifische qualitative und quantitative metabolische Modelle mit entsprechenden Modellierungsformalismen analysiert.

Anschließend wird die integrative Analyse metabolischer Modelle der Kulturpflanzen Kartoffel und Mais durchgeführt. Die Erkenntnisse und Resultate der vorliegenden Arbeit werden abschließend in Kapitel 8 zusammengefasst und diskutiert. Es wird ein Ausblick über zukünftige Anwendungen und Weiterentwicklungen gegeben.



**Abbildung 1.3:** Struktur der vorliegenden Arbeit.





## 2 Grundlagen

Dieses Kapitel erklärt grundlegende Begriffe, die bei der metabolischen Modellierung Verwendung finden. Zu Beginn werden Begriffe aus der Biochemie erläutert und eine Einführung in die verwendete Mathematik und Informatik gegeben. In den Grundlagen der Systembiologie wird die Synergie zwischen Biochemie, Mathematik und Informatik beschrieben. Das Vorgehen in diesem Kapitel ist in Abbildung 2.1 dargestellt.

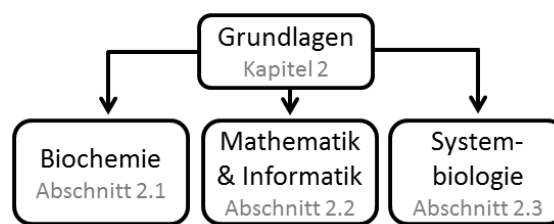


Abbildung 2.1: Struktur des Kapitels Grundlagen.

### 2.1 Grundlagen aus der Biochemie

Dieser Abschnitt befasst sich mit den biochemischen Grundlagen zur metabolischen Modellierung. Der zentrale Begriff Metabolismus, wird basierend auf Inhalten des Buches *Biochemie* [11] Kapitel 15, eingeführt. Die erste schriftliche Abhandlung zum Thema Biochemie verfasste *Vinzenz Kletzinsky* im Jahr 1858, mit seinem *Compendium der Biochemie* [87]. In diesem Buch wird in Kapitel 2 die Chemie der biochemischen Prozesse erörtert, die allgegenwärtig in lebenden Zellen und an allen zellulären Funktionen wesentlich beteiligt sind.

#### 2.1.1 Merkmale von Lebewesen

Lebewesen werden in verschiedene Domänen unterteilt. Es gibt *Prokaryoten* (Bakterien, Archea), die keinen Zellkern besitzen und *Eukaryoten*, die einen Zellkern besitzen. Zudem enthalten Eukaryoten durch Membranen voneinander abgegrenzte Reaktionsräume, die *Kompartimente* genannt werden. Lebewesen sind somit aus stark organisierten Zellen aufgebaut, die Gewebe, Organe und letztlich einen Organismus bilden. Die Grundbausteine von Zellen sind *Biomoleküle* (einfache oder komplexe chemische Verbindungen), die als Informationsträger (DNA und RNA), Energielieferanten und Speicher (Kohlehydrate, Proteine, Fette) wichtige Funktionen der Zelle übernehmen. Diese Biomoleküle können aufgenommen, transportiert, synthetisiert, umgewandelt, degradiert und abgegeben werden. Die Gesamtheit dieser chemischen Vorgänge wird *Metabolismus* (siehe Abschnitt 2.1.4) genannt und findet in einem Netzwerk *biochemischer Reaktionen* statt, die das Wachstum und die Energieversorgung der Zelle erhalten.

### 2.1.2 Biochemische Reaktion

Biochemische Reaktionen werden in *enzymkatalysierte* und *spontane* Reaktionen sowie *Transportreaktionen* unterteilt.

**Enzymkatalysierte Reaktionen** Die meisten Reaktionen laufen langsam ab und benötigen Katalysatoren. Enzyme sind Katalysatoren, die die Aktivierungsenergie für eine Reaktion verringern und die Reaktionsrate steigern. Dabei werden *Metabolite* als *Substrate* in das substratspezifische, aktive Zentrum des Enzyms gebunden, der Enzym-Substrat-Komplex entsteht. Das Substrat wird in ein *Produkt* umgewandelt und verlässt das Enzym, sodass das Enzym eine neue Reaktion katalysieren kann. Entsprechend der von ihnen katalysierten Reaktion werden Enzyme mittels *EC-Nummern* (engl. Enzyme Commission numbers) codiert. Die EC-Nummer besteht aus vier durch Punkte getrennt Zahlen, beispielsweise steht die EC-Nummer 3.2.1.20 für das Enzym  $\alpha$ -Glykosidase. Die Zahlen legen Enzymklassen fest, wobei sechs Hauptklassen unterschieden werden (erste Stelle kann die Zahlen 1-6 annehmen) und die weiteren drei Stellen eine Spezifizierung der Reaktion angeben (Details siehe [82]). Viele Enzyme benötigen Cofaktoren (Metallionen, kleine organische Moleküle) um aktiv zu sein. Es gibt auch Inhibitoren (Ionen, kleine organische Moleküle), die Enzyme binden und die Aktivität hemmen. Diese Inhibition kann reversibel oder irreversibel, kompetitiv oder nicht-kompetitiv sein. Auch Produkte oder Substrate einer Reaktion können inhibierend wirken (Produkt- oder Substratinhibition). Enzyme besitzen des Weiteren kinetische Eigenschaften, die in Kapitel 3 Abschnitt 3.2, eingeführt werden.

**Spontane Reaktionen** Eine spontane Reaktion läuft von selbst ab, da diese eine sehr geringe Aktivierungsenergie benötigt, die bereits durch die Umgebungstemperatur geliefert wird.

**Transportreaktionen** Membranen sind für Ionen und polare Moleküle impermeabel. Transportproteine befinden sich in Membranen und transportieren Moleküle passiv, entlang eines Konzentrationsgefälles oder unter Energieverbrauch aktiv, gegen ein Konzentrationsgefälle. Transportreaktionen können Uniport (einzelne Moleküle in eine Richtung), Symport (zwei Moleküle gleichzeitig in eine Richtung) oder Antiport (zwei Moleküle gleichzeitig in entgegengesetzte Richtungen) erfolgen. Ebenso wie Enzyme können Transportproteine inhibiert werden und besitzen kinetische Eigenschaften.

**Formale Schreibweise** Eine biochemische Reaktion wird formal als *Reaktionsgleichung* geschrieben, sodass die Substrate einer Reaktion links und die Produkte rechts des Reaktionspfeils stehen. Zum Beispiel wird die Umwandlung von Maltose in zwei Moleküle Glukose, durch das Enzym  $\alpha$ -Glykosidase (EC: 3.2.1.20), katalysiert:  $Maltose \rightarrow 2 Glukose$ . Quantitative Beziehungen zwischen Substraten und Produkten, in einer ausgeglichenen Reaktion, werden durch *Stöchiometrien* repräsentiert. In der Beispielreaktion ist die 2 vor Glukose ein stöchiometrischer Koeffizient. Hingegen wird eine Stöchiometrie von 1 nicht explizit angegeben, sodass vor Maltose keine 1 geschrieben steht. Des Weiteren wird in dieser Schreibweise die *Reaktionsrichtung* durch Reaktionspfeile gekennzeichnet. Eine *irreversible* Reaktion ist durch einen Reaktionspfeil, der in eine Richtung weist, dargestellt (wie im Beispiel). Der Reaktionspfeil einer *reversiblen* Reaktion ist ein Gleichgewichtspfeil

↔. Diese Reaktion besteht aus einer *Hinreaktion*, die von Substraten zu Produkten verläuft und einer *Rückreaktion* in umgekehrter Richtung.

### 2.1.3 Stoffwechselwege

Metabolite können an verschiedenen Reaktionen teilnehmen. Produkte einer Reaktion gehen als Substrate in andere Reaktionen ein. Diese zusammenhängenden Reaktionen werden abhängig von deren metabolischer Funktion *Stoffwechselwegen* (engl. pathways) zugeordnet, z. B. der Glykolyse. Stoffwechselwege sind für die *Biosynthese* und *Degradation* chemischer Verbindungen unerlässlich. *Intermediäre* Metabolite verbinden Stoffwechselwege miteinander. Transportprozesse ermöglichen den Austausch von Metaboliten über Kompartimente hinweg. Stoffaustausch mit der Umgebung einer Zelle ergibt eine Unterscheidung von Metaboliten, die sich innerhalb der Zelle befinden (*interne* Metabolite) und denen, die sich außerhalb der Zelle befinden (*externe* Metabolite). Aufgrund der Verbindungen verschiedener Stoffwechselwege entsteht ein komplexes metabolisches Netzwerk, der *Metabolismus*.

### 2.1.4 Metabolismus

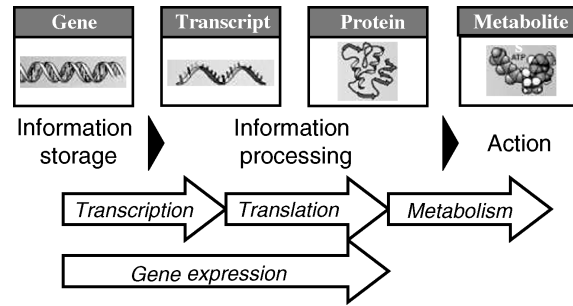
Im Metabolismus werden zwei Phasen unterschieden, der *Katabolismus* und der *Anabolismus*. Im Katabolismus wird durch Degradation von komplexen Verbindungen (Proteine, Kohlehydrate, Lipide) zu einfachen Verbindungen (Wasser, Kohlenstoffdioxid), Energie gewonnen. Die Energie wird für energieverbrauchende, zelluläre Prozesse verwendet und die einfachen Verbindungen bilden die Substrate zur Biosynthese im Anabolismus. Der Anabolismus ist ein energieverbrauchender Prozess zur Biosynthese komplexer Verbindungen.

Weiterhin gibt es eine Unterteilung in *Primär- und Sekundärmetabolismus*. Der Primärmetabolismus besteht aus wesentlichen Prozessen, die unentbehrlich und lebensnotwendig sind, die das Wachstum und die Entwicklung eines Organismus ermöglichen. Diese wesentlichen Prozesse stimmen in den meisten Organismen weitgehend überein. Es handelt sich um ein komplexes Netzwerk, bestehend aus Protein-, Kohlehydrat-, Lipid-, Energie- und Nukleinsäurestoffwechsel. Dabei bilden die Glykolyse, der Citrat-Zyklus, die oxidative Phosphorylierung, der oxidative Pentosephosphatweg und die Fettsäuresynthese die zentralen metabolischen Stoffwechselwege (nähere Erläuterungen siehe [11]). Der Primärmetabolismus liefert Substrate für den Sekundärmetabolismus, welcher Verbindungen synthetisiert, die für die meisten Organismen nicht essentiell sind. Der Sekundärmetabolismus zeichnet sich durch seine Vielfalt und Adaptivität aus.

### 2.1.5 Beeinflussung des Metabolismus

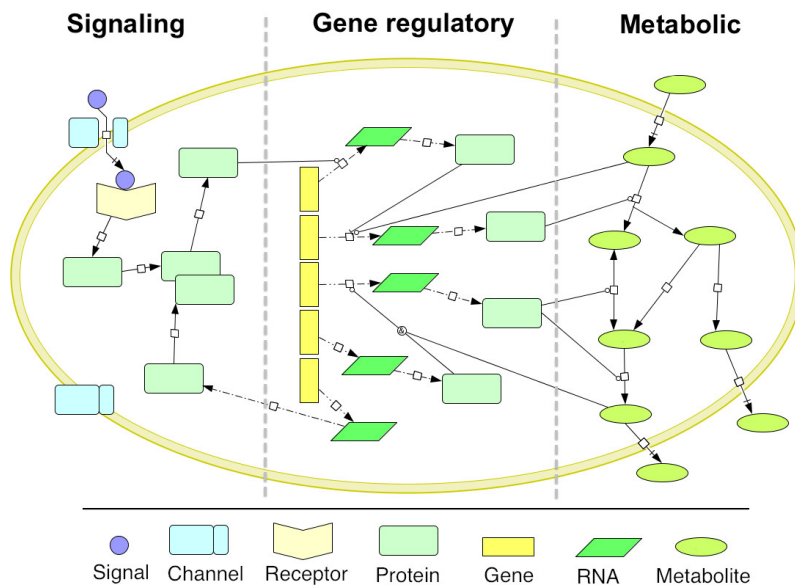
Der Informationsfluss und somit die Verknüpfung des Metabolismus mit dessen Regulation erfolgt nach dem Zentralen Dogma der Molekularbiologie [29, 30], siehe Abbildung 2.2.

Proteine (z. B. Enzyme) werden von den codierenden Bereichen der DNA, den *Genen*, synthetisiert. Dieser Prozess wird *Genexpression* genannt und besteht aus *Transkription* (DNA → RNA) und *Translation* (RNA → Protein). Gene können durch andere Proteine, den *Transkriptionsfaktoren*, aktiviert (Aktivator) oder reprimiert (Repressor) werden. Es handelt sich um die *Genregulation*. Auch die Transkriptionsfaktoren werden in der DNA codiert und werden der Regulation unterzogen, sodass ein komplexes *Genregulationsnetzwerk* auf den Metabolismus Einfluss nimmt.



**Abbildung 2.2:** Informationsfluss in der Zelle von Genen zu Metaboliten, entnommen aus [79].

Zellen sind zudem in der Lage, auf Veränderungen innerhalb und außerhalb zu reagieren. Die Bindung von *Signalmolekülen* (Hormone, Wachstumsfaktoren) an *Rezeptoren* setzt eine *Signaltransduktionskaskade* in Gang. Diese Kaskade beinhaltet Botenstoffe und Enzyme, die als molekulare Schalter, biochemische Reaktionen beeinflussen. Solche Signaltransduktionkaskaden stehen untereinander in Verbindung und bilden ein *Signaltransduktionsnetzwerk*, welches ebenso wie das Genregulationsnetzwerk in Verbindung mit dem Metabolismus steht, siehe Abbildung 2.3.



**Abbildung 2.3:** Zusammenhang zwischen Signaltransduktion, Genregulation und Metabolismus, entnommen aus [106].

## 2.2 Grundlagen aus der Mathematik und Informatik

Im Folgenden werden die zur Modellierung metabolischer Modelle notwendigen Grundlagen der Mathematik und Informatik erläutert. Zum Verständnis der Arbeit werden Begriffe der Mengenlehre, Logik und Graphentheorie vorgestellt, die auf Definitionen aus Junker und Schreiber [79], Seite 16 bis 22, Jungnickel [76] und Goos [58] basieren.

### 2.2.1 Mengen und Mengenoperationen

Georg Cantor wird heute im Allgemeinen als Begründer der Mengenlehre angesehen. Die Definition einer *Menge* hat Cantor wie folgt formuliert [21]:

„Unter einer *Menge* verstehen wir jede Zusammenfassung  $M$  von bestimmten wohlunterschiedenen Objekten  $m$  unserer Anschauung oder unseres Denkens (welche die *Elemente* von  $M$  genannt werden) zu einem Ganzen.“

Der Ausdruck „ $m$  ist Element der Menge  $M$ “, wird  $m \in M$  geschrieben und als Element-Relation bezeichnet. Ist ein Element  $x$  nicht in der Menge  $M$ , wird  $x \notin M$  geschrieben. Es gibt verschiedene Schreibweisen zur Charakterisierung von Mengen. Als Beispiel dienen die Elemente 1, 2 und 3 der Menge  $M$ , wobei die Menge in der aufzählenden Schreibweise (extensional) als  $M = \{1, 2, 3\}$  und in der beschreibenden Schreibweise (intensional) als  $M = \{m \in \mathbb{N} \mid 0 < m < 4\}$  charakterisiert wird. In Worten ausgedrückt: „ $m$  ist Element der Menge der natürlichen Zahlen ( $\mathbb{N}$ ) mit der Eigenschaft, dass  $m$  die Zahlen zwischen 0 und 4 annimmt“.

Eine Menge, die keine Elemente enthält, heißt *leere Menge*,  $\emptyset$ . Eine Menge  $M$  heißt *Teilmenge* einer Menge  $O$ , geschrieben  $M \subseteq O$ , wenn aus  $x \in M$  stets  $x \in O$  folgt, also alle Elemente  $x$  der Menge  $M$  in  $O$  enthalten sind. Die Mengen  $M$  und  $O$  sind genau dann gleich, geschrieben  $M = O$ , wenn  $M \subseteq O$  und  $O \subseteq M$ , sonst ist  $M$  ungleich  $O$ , geschrieben  $M \neq O$ .  $M$  heißt genau dann *echte Teilmenge* von  $O$ , geschrieben  $M \subset O$ , wenn  $M \subseteq O$  und  $M \neq O$  ist.

Weitere Mengen-Operationen sind wie folgt definiert. Die *Vereinigung* zweier Mengen ist eine Menge, die entweder aus allen Elementen aus  $M$  oder  $O$  besteht, geschrieben  $M \cup O = \{x \mid (x \in M) \text{ oder } (x \in O)\}$ . Der *Schnitt* zweier Mengen ist eine Menge, die aus Elementen beider Mengen besteht, geschrieben  $M \cap O = \{x \mid (x \in M) \text{ und } (x \in O)\}$ . Haben die beiden Mengen kein gemeinsames Element, heißen die Mengen *disjunkt*, geschrieben  $M \cap O = \emptyset$ . Das *Komplement* einer Menge ist eine Menge in der keine Elemente aus  $M$  enthalten sind, geschrieben  $\overline{M} = \{x \in G \mid x \notin M\}$  (bezogen auf die Grundmenge  $G$ ). Die *Differenz* zweier Mengen ist eine Menge die alle Elemente aus  $M$  enthält, die nicht Element  $O$  sind, geschrieben  $M \setminus O = M \cap \overline{O}$ . Das *kartesische Produkt* zweier Mengen ist die Menge aller *geordneten Paare*  $(x, y)$ , wobei die erste Komponente des Paares ein Element  $x$  der ersten Menge  $M$  und die zweite Komponente des Paares ein Element  $y$  der zweiten Menge  $O$  bezeichnet, geschrieben  $M \times O = \{(x, y) \mid (x \in M) \text{ und } (y \in O)\}$ .

### 2.2.2 Relationen und Funktionen

Eine (*zweistellige*) *Relation*  $R$  zwischen zwei Mengen  $M$  und  $O$  ist eine Teilmenge des kartesischen Produkts  $M \times O$ , geschrieben  $R \subseteq M \times O$ . Ist  $x \in M$ ,  $y \in O$  und gehört das geordnete Paar  $(x, y)$  zur Relation  $R$ , geschrieben  $(x, y) \in R$ , heißt es „ $x$  steht in Relation zu  $y$ “. Steht  $x$  nicht in Relation zu  $y$ , wird  $(x, y) \notin R$  geschrieben. Als zweistellige Relation wird auch der Spezialfall  $M = M$ , also  $R \subseteq M \times M$  bezeichnet. Relationen gibt es auch als *n-stellige Relation*  $R$  zwischen mehr als zwei Mengen, geschrieben  $R \subseteq M_1 \times M_2 \times \dots \times M_n$ .

Eine (*partielle*) *Funktion* ist eine spezielle zweistellige Relation  $f \subseteq M \times O$ , die *rechtseindeutig* ist, geschrieben  $f : M \rightarrow O$ . Ein Element  $x \in M$  steht mit höchstens einem Element  $y \in O$  in Relation, sodass aus  $(x, y) \in f$  und  $(x, z) \in f$  folgt  $y = z$ . Die Menge  $M$  wird *Definitionsbereich* der Funktion und  $o$  wird *Wertebereich* der Funktion genannt. Anstatt  $(x, y) \in f$  wird  $f(x) = y$  geschrieben und  $x$  wird in diesem Zusammenhang

*Urbild von  $y$  unter  $f$*  genannt. Ist  $f$  zusätzlich linkstotal, d. h. jedes Element  $x \in M$  steht mit genau einem Element  $y \in O$  in Relation, sodass es kein Element  $x \in M$  gibt, für das  $f(x)$  nicht definiert ist, wird *totale Funktion* oder *Abbildung* genannt. Jede totale Funktion ist auch eine partielle Funktion. Eine totale Funktion  $f : M \rightarrow O$  heißt *injektiv*, wenn es zu jedem  $y \in O$  höchstens ein  $x \in M$  mit  $y = f(x)$  gibt, d. h.  $f$  ist *linkseindeutig*. Die Funktion  $f$  heißt *surjektiv*, wenn  $f(M) = O$ , d. h.  $f$  ist *rechtstotal*. Wenn  $f$  zugleich injektiv und surjektiv ist, wird die Funktion *bijektiv* bezeichnet. Zu jeder bijektiven Funktion  $f : M \rightarrow O$  existiert eine *Umkehrfunktion* von  $f$ , geschrieben  $\bar{f} : O \rightarrow M$ , d. h.  $\bar{f}$  bildet ein Element  $y \in O$  auf das eindeutig bestimmte Urbild von  $y$  (also  $x$ ) unter  $f$  ab, geschrieben  $\bar{f}(y) = x$ .

## 2.2.3 Logik

Die Lehre des vernünftigen Schlussfolgerns wird Logik genannt und beinhaltet u. a. die Teilgebiete *Aussagenlogik* und *Prädikatenlogik*.

### 2.2.3.1 Aussagenlogik

In der auf die Stoiker und Aristoteles zurückgehenden *Aussagenlogik* werden Aussagen und deren logische Verknüpfungen untersucht. Der griechische Philosoph Aristoteles schrieb bereits in seinem Werk, *Analytica priora*, 350 v. Chr. [6]:

„Eine *Aussage* ist ein sprachliches Gebilde, von dem es sinnvoll ist zu sagen, es sei *wahr* oder *falsch*.“

Es werden also nur Aussagen betrachtet, die entweder *wahr* (1) oder *falsch* (0) sind, d. h. jeder Aussage ist ein *Wahrheitswert* zugeordnet. Die Aussage „es regnet“, kann eindeutig wahr oder falsch annehmen. Eine zusammengesetzte Aussage besteht aus Teilaussagen und *Junktoren*. Ist eine Aussage nicht weiter in Teilaussagen zerlegbar, wird diese *atomare Aussage* genannt. Atomare Aussagen werden mittels *Junktoren* und  $(\wedge)$ , *oder*  $(\vee)$ , *nicht*  $(\neg)$ , *impliziert*  $(\Rightarrow)$ , *äquivalent*  $(\Leftrightarrow)$ , *entweder-oder*  $(\dot{\vee})$  verknüpft. Aussagen werden als *aussagenlogische Formeln* notiert, indem eine (atomare oder zusammengesetzte) Aussage durch Buchstaben abstrahiert wird. Die zusammengesetzte Aussage „es regnet und die Wiese ist nass“, wird durch die Formel  $A \wedge B$  abstrahiert.

Mit Hilfe von *Wahrheitstabellen* ist das Ergebnis einer Aussage unter allen möglichen Belegungen der Variablen bestimmbar.

### 2.2.3.2 Prädikatenlogik

Aufgrund der begrenzten Ausdrucksstärke von Aussagenlogik (zu starke Abstraktion), wird im Folgenden die Prädikatenlogik vorgestellt. Gottlob Frege führte in seiner *Begriffsschrift* [59] von 1879 mit dem Untertitel, „Eine der arithmetischen nachgebildete Formelsprache des reinen Denkens“, die Prädikatenlogik ein. In der Aussagenlogik werden atomare Aussagen, die *als Ganzes* wahr oder falsch sind, betrachtet. In der Prädikatenlogik werden *parametrisierte Aussagen* verwendet, die durch das Einsetzen eines Elementes aus einer Menge  $M$ , zu wahren oder falschen Aussagen werden. Eine parametrisierte Aussage heißt *Prädikat* und die Menge  $M$  wird als *Grundmenge* bezeichnet. Beispielsweise kann das Prädikat, „*ist Primzahl*( $x$ )“, wobei  $\mathbb{N}$  die Grundmenge sein soll, eindeutig wahr oder falsch annehmen. Prädikate können auch *Relationen* sein die mehrere Elemente aus einer oder mehreren Grundmengen miteinander verknüpfen, z. B., „*kleiner als*( $x, y$ )“, wobei  $\mathbb{N}$  die

Grundmenge sein soll. Enthält ein Prädikat  $n$  Variablen, wird es *n-stelliges Prädikat* genannt. Es gibt atomare Prädikate, die durch Junktoren (wie in Abschnitt 2.2.3.1 erläutert) zu zusammengesetzten Prädikaten verknüpft werden können. Ein zusammengesetztes Prädikat ist z. B., „*ist\_Primzahl(x) ∧ kleiner\_als(x, y)*“, wobei  $\mathbb{N}$  die Grundmenge sein soll.

Wesentlicher Teil der Prädikatenlogik sind neben Prädikaten die *Quantoren*. Ist  $P(x)$  ein Prädikat, wird für die Aussage „ $P(x)$  ist für alle  $x$  aus der Grundmenge wahr“,  $\forall x : P(x)$  geschrieben und  $\forall$  wird als *Allquantor* bezeichnet. Für die Aussage, „Es existiert (mindestens) ein  $x$  aus der Grundmenge für das  $P(x)$  wahr ist“, wird  $\exists x : P(x)$  geschrieben und  $\exists$  wird als *Existenzquantor* bezeichnet. In einem Ausdruck der Art  $\exists x \forall y : P(x, y, z)$ , heißen  $x$  und  $y$  gebundene Variablen und  $z$  freie Variable. Werden Quantoren geschachtelt, ist die Reihenfolge wichtig. Deshalb besteht ein Unterschied zwischen  $\forall x \exists y : y > x$  und  $\exists y \forall x : y > x$ , wobei  $\mathbb{N}$  die Grundmenge sein soll.

Diese Einführung in die Aussagen- und Prädikatenlogik legt die Grundlage für die im folgenden Abschnitt und in der Methodik (Kapitel 5) verwendete prädikatenlogische Formalisierung von Graphen. Für detaillierte Definitionen und Erklärungen von z. B. Syntax und Semantik sei auf die entsprechende Literatur [58] verwiesen.

## 2.2.4 Graphen

Die Struktur eines metabolischen Modells ist ein Netzwerk, welches Objekte zueinander in Beziehung setzt und dessen mathematische Entsprechung ein Graph ist. Ein Graph besteht aus einer definierten Menge von Knoten, die über eine definierte Menge von Kanten miteinander verbunden sind.

**Ungerichteter Graph** Ein *ungerichteter Graph*  $G$  besteht aus einer endlichen, nicht-leeren Menge  $V$ , die *Knotenmenge* von  $G$  genannt wird und einer endlichen, nicht-leeren Menge  $E \subseteq \{\{x, y\} | x \in V \wedge y \in V\}$ , die *Kantenmenge* von  $G$  genannt wird. Die Elemente aus  $V$  heißen *Knoten* (engl. vertices) und die Elemente aus  $E$  heißen *Kanten* von  $G$  (engl. edges), hier speziell ungerichtete Kanten. Jede Kante  $e \in E$  wird als ein ungeordnetes Paar von Knoten  $e = \{x, y\}$  definiert. Die Knoten  $x$  und  $y$  heißen *inzident* mit der Kante  $e$  und *adjazent* zueinander. Ungerichtete Graphen werden zum Beispiel zur Modellierung von Protein-Protein Interaktionsnetzwerken verwendet.

**Gerichteter Graph** Im Gegensatz zu einem ungerichteten Graphen, wird in einem *gerichteten Graphen*  $G = (V, E)$ , jede Kante  $e \in E$  als ein geordnetes Paar von Knoten  $e = (x, y)$  definiert:  $E \subseteq \{(x, y) : x \in V \wedge y \in V\}$ . Knoten  $x$  heißt *Ausgangsknoten* von  $e$  und Knoten  $y$  wird als *Endknoten* von  $e$  bezeichnet. Kante  $e$  verläuft somit von  $x$  nach  $y$ . Gerichtete Graphen werden zum Beispiel zur Modellierung von Genregulationsnetzwerken verwendet.

**Darstellung von Graphen** Die Darstellung von Graphen kann: (1) *abstrakt*, durch die Angabe von Knotenmengen  $V$  und Kantenmengen  $E$  erfolgen, (2) *graphisch*, wobei Knoten als Punkte und Kanten als Verbindungslinien zwischen Punkten dargestellt werden. Die Kanten von gerichteten Graphen, z. B.  $e = (x, y)$  wird als Pfeil von Knoten  $x$  nach Knoten  $y$  dargestellt. Die Position von Knoten und die Zeichnung der Kanten werden *Layout* des Graphen genannt, welches für gleiche Graphen unterschiedlich sein kann. (3) *Adjazenzliste/-matrix*, wobei die Adjazenzliste zu jedem Knoten  $x$  eine Liste aller Knoten

angibt, zu denen eine von  $x$  ausgehende Kante führt. Die Adjazenzmatrix  $A$  mit den Einträgen  $a_{i,j}$  ist eine  $n \times n$ -Matrix, deren Zeilen und Spalten mit Knoten beschriftet sind. Die Einträge  $a_{i,j}$  sind entweder 0 oder 1. Wenn es eine Kante von Knoten  $x$  zu Knoten  $y$  gibt, wird in der mit Knoten  $x$  beschrifteten Zeile und der mit Knoten  $y$  beschrifteten Spalte der Eintrag 1 erfolgen:  $a_{i,j} = 1$  falls  $(v_i, v_j) \in E$ . Existiert keine Kante zwischen diesen Knoten wird der Eintrag 0 vergeben:  $a_{i,j} = 0$  falls  $(v_i, v_j) \notin E$ .

**Wege und Zyklen in Graphen** In einem Graph  $G = (V, E)$  heißt eine Folge von Knoten  $v_1, v_2, \dots, v_n$  mit  $v_i \in V$  und  $i = 1, \dots, n$  *Weg* von  $v_1$  nach  $v_n$  wenn gilt:  $\forall 1 \leq i < n : (v_i, v_{i+1}) \in E$ . Ein *Zyklus* ist ein Weg, bei dem Start- und Endknoten gleich sind:  $v_1 = v_n$ .

**Ähnlichkeit von Graphen** Es kann überprüft werden, ob von zwei (gerichteten oder ungerichteten) Graphen  $G = (V, E)$  und  $G' = (V', E')$ ,  $G'$  *Teilgraph* von  $G$  ist.  $G'$  heißt *Teilgraph* von  $G$ , falls  $V' \subseteq V$  und  $E' \subseteq E$  ist.  $G'$  heißt *induzierter Teilgraph* von  $G$ , falls  $V' \subseteq V$  und  $E \subseteq \{e \in E : \text{die mit } e \text{ inzidenten Knoten liegen in } V'\}$ . Zwei (gerichteten oder ungerichteten) Graphen  $G = (V, E)$  und  $G' = (V', E')$  sind *gleich*, geschrieben  $G = G'$ , falls beide dieselbe Knoten- und Kantenmenge besitzen, d. h.  $G = G' \Leftrightarrow V = V' \wedge E = E'$ . Die *Isomorphie* zweier (gerichteter oder ungerichteter) Graphen  $G = (V, E)$  und  $G' = (V', E')$ , geschrieben  $G \cong G'$  (in Worten  $G$  ist *isomorph* zu  $G'$ ), falls eine bijektive Abbildung  $f : V \rightarrow V'$  existiert, sodass für alle Knoten  $x \in V$  und  $y \in V'$  gilt, (1) für gerichtete Graphen:  $(x, y) \in E \Leftrightarrow (f(x), f(y)) \in E'$ , (2) für ungerichtete Graphen:  $\{x, y\} \in E \Leftrightarrow \{f(x), f(y)\} \in E'$ .

**Attributierte Graphen** Ein (gerichteter oder ungerichteter) Graph  $G$  heißt *attribuierter Graph*, wenn dieser Attribute mittels Abbildungen von Kanten- und/oder Knotenmengen zu geeigneten Mengen vergibt. Der attributierte Graph  $G = (V, E, A)$  besitzt eine Attributmenge  $A = \{m_v, m_e\}$ , wobei die folgende *Knotenabbildung*  $m_v : V \rightarrow G_v$  und die *Kantenabbildung*  $m_e : E \rightarrow G_e$  den Knoten und Kanten Attribute aus geeigneten Mengen  $G_v, G_e$  zuordnen. In biologischen Netzwerken wird häufig eine Knotenabbildung verwendet, die jedem Knoten eine Beschriftung und eine Kantenabbildung, die jeder Kante ein Gewicht zuordnet. Der attributierte Graph  $G = (V, E, A)$  mit  $A = \{l, g\}$  enthält die Knotenabbildung  $l : V \rightarrow \Sigma^*$  und vergibt jedem Knoten  $v \in V$  ein Wort über dem Alphabet  $\Sigma$ , die Beschriftung  $l(v)$ . Die Kantenabbildung  $w : E \rightarrow \mathbb{R}_{\geq 0}$  vergibt jeder Kante  $e \in E$  eine reelle Zahl größer als 0, das Gewicht  $w(e)$ . Attributierte (gerichtete oder ungerichtete) Graphen werden verwendet wenn die Modellierung biologischer Netzwerke erfordert, Knoten und Kanten eines Graphen weitere Informationen zuzuordnen.

**Bipartite Graphen** Ein (gerichteter oder ungerichteter) *bipartiter* Graph  $G = (V, E)$  besitzt eine Knotenmenge  $V = U \cup W$ , die in zwei nicht-leere, disjunkte Teilmengen  $U$  und  $W$  zerfällt, geschrieben  $U \cap W = \emptyset$ , und wenn jede Kante  $e \in E$  einen Knoten  $x \in U$  und einen Knoten  $y \in W$  abbildet.

Nach [91, 170] entsteht durch die Projektion auf eine der beiden Knotenmengen aus einem *bipartiten* Graphen ein *zugehöriger unipartiter* Graph für die entsprechende Knotenmenge.

**Zugehöriger unipartiter Graph** Sei  $G = (V, E)$  ein gerichteter, *bipartiter* Graph mit der Knotenmenge  $V = U \cup W$ , dann ist der Graph  $G_U = (U, E_U)$  mit der Kantenmenge  $E_U = \{(x, z) | (x, y) \in E \wedge (y, z) \in E \wedge x, z \in U \wedge y \in W\}$ , der zu  $G$  und Knotenmenge  $U$



zugehörige gerichtete, *unipartite* Graph. Im Fall eines ungerichteten, *bipartiten* Graphen erfolgt die Kantendefinition des zugehörigen ungerichteten, *unipartiten* Graphen nach  $E_U = \{\{x, z\} | \{x, y\} \in E \wedge \{y, z\} \in E \wedge x, z \in U \wedge y \in W\}$ . Der zugehörige *unipartite* Graph enthält somit nur eine der beiden Knotenmengen des *bipartiten* Graphen. Zwei Knoten sind genau dann verbunden, wenn diese im *bipartiten* Graphen mit einem gemeinsamen Nachbarn aus der anderen Knotenmenge verbunden sind.

**Schlichter Graph** Ein *schlichter* Graph enthält weder *Mehrfachkanten*, d. h. zwischen zwei Knoten existiert nur eine Kante, noch *Schlingen*, d. h. ein Knoten besitzt keine Kante zu sich selbst. Wird die Kantenmenge eines ungerichteten Graphen  $G = (V, E)$  wie folgt geschrieben  $E \subseteq \{\{x, y\} | x \in V \wedge y \in V \wedge (x \neq y)\}$ , existieren keine Mehrfachkanten und Schlingen. Bei einem gerichteten Graphen  $G = (V, E)$  verhindert die Kantendefinition  $E \subseteq \{(x, y) | x \in V \wedge y \in V \wedge (x \neq y)\}$  Schlingen und die zusätzliche Bedingung  $\forall x, y \in V : (x, y) \in E \Rightarrow (y, x) \notin E$  verhindert Mehrfachkanten. Ein ungerichteter, bipartiter Graph ist aufgrund der ungeordneten Kantenmenge bereits schlicht. Ein gerichteter, bipartiter Graph ist schlicht, wenn die Bedingung  $\forall x, x' \in U, \forall y \in W : (x, y) \in E \wedge (y, x') \in E \Rightarrow (x \neq x')$  gilt. Im Folgenden sind alle Graphen frei von Schlingen, auch wenn die Bezeichnung nicht explizit verwendet wird.

Die Graphentheorie beinhaltet weitere Graphen wie z. B. *Hypergraphen*, *Multigraphen* und *Bäume*, die für das Verständnis dieser Arbeit nicht notwendig sind. Für weitere Definitionen und Erklärungen sei auf die entsprechende Literatur, z. B. Diestel *Graph Theory* [39], verwiesen.

## 2.3 Grundlagen aus der Systembiologie

Das letzte Jahrhundert erbrachte enorme Fortschritte in den Gebieten Biochemie und Molekularbiologie. Die Komponenten des Lebens wurden aus Zellen extrahiert, identifiziert und ihre Eigenschaften bestimmt. Die Informationen des Genoms, Transkriptoms, Proteoms und Metaboloms geben nur einen kleinen Einblick in die Dynamik einer lebenden Zelle. Das Gebiet der Systembiologie nutzt dieses Wissen und verbindet die Biologie, Mathematik und Informatik, um die Mechanismen biologischer Vorgänge integriert zu untersuchen.

Nach Hiroaki Kitano [84, 85], der 2002 den Begriff Systembiologie prägte, leitet sich das Verständnis eines biologischen Systems auf Systemebene von den Kenntnissen über die Systemstruktur und Systemdynamik ab. Zur Beschreibung eines biologischen Systems nutzt die Systembiologie die mathematische Modellierung. Handelt es sich um die Charakterisierung des Metabolismus, werden biochemische Reaktionen auf Modelle abgebildet und analysiert. Dieser Prozess wird *metabolische Modellierung* genannt.

### 2.3.1 Metabolische Modellierung

Die Modellierung basiert auf einer subjektiven und selektiven Prozedur, sodass ein Modell nur einen spezifischen Aspekt der Realität repräsentiert.

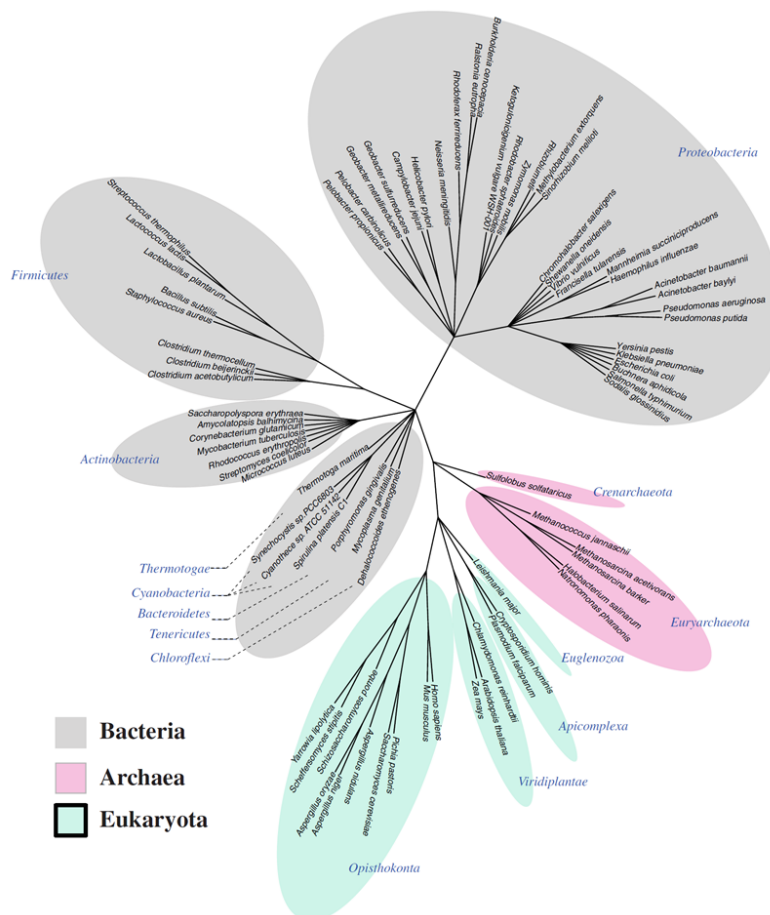
#### 2.3.1.1 Metabolisches Modell

Ein *metabolisches Modell* ist ein Abbild des Metabolismus und beschreibt einen Ausschnitt eines realen biologischen Systems.

Das dem metabolischen Modell zugrunde liegende Netzwerk zusammenhängender, biochemischer Reaktionen wird strukturell in Form eines gerichteten bipartiten Graphen

(siehe Abschnitt 2.2.4) beschrieben. Die biochemischen Reaktionen bilden die eine und Metabolite die andere Knotenmenge. Die gerichteten Kanten führen von Metaboliten zu Reaktionen und bilden somit die Substrate einer Reaktion, die verbraucht werden. Umgekehrt führen gerichtete Kanten von Reaktionen zu Metaboliten und bilden somit Produkte einer Reaktion, die produziert werden. Die Stöchiometrie beschreibt die quantitativen Beziehungen zwischen Substraten und Produkten einer Reaktion. Reaktionen können irreversibel (Kanten führen in eine Richtung) und reversibel (Kanten führen in beide Richtungen und bilden somit eine Hin- und Rückreaktion ab) sein. Wird die subzelluläre Lokalisation der Reaktionen (Kompartimente) modelliert, ermöglichen intrazelluläre Transportprozesse den Austausch von Metaboliten. Zur Abgrenzung von der Umgebung werden extrazelluläre Transportprozesse, Import- und Exportreaktionen verwendet. Importreaktionen importieren Metaboliten aus der Umgebung und Exportreaktionen exportieren Metaboliten in die Umgebung.

Metabolische Modelle sind derzeit die am besten biochemisch, kinetisch und thermodynamisch charakterisierten biologischen Netzwerke [117]. Einen Einblick in die Vielfalt der Organismen, die bisher modelliert wurden, zeigt der phylogenetische Baum aus der Arbeit von Xu und Koautoren [169], siehe Abbildung 2.4.



**Abbildung 2.4:** Phylogenetischer Baum metabolischer Modelle der drei Domänen Bakterien (Grau), Achea (Magenta) und Eukaryoten (Grün), entnommen aus [169].

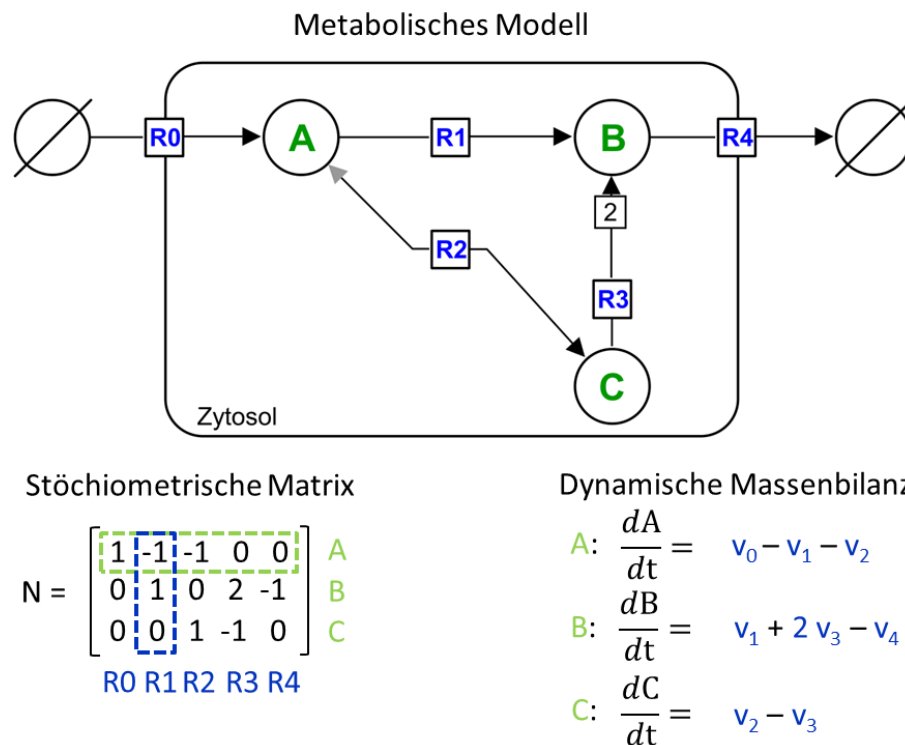
Farblich wird in der Abbildung zwischen den drei Domänen Bakterien (54), Archea (6) und Eukaryoten (16) unterschieden. In Klammern ist die Anzahl der Modelle pro Domäne

angegeben. Diese Modelle unterscheiden sich in deren Größe und enthaltenen Details. Allen gemein ist, dass ein metabolisches Modell aus einem Netzwerk biochemischer Reaktionen besteht, die eine Menge von Metaboliten ineinander umwandeln.

### 2.3.1.2 Rekonstruktion metabolischer Modelle

Die Erstellung (Rekonstruktion) metabolischer Modelle ist durch zwei verschiedene Ansätze möglich, dem *Top-down* und dem *Bottom-up* Ansatz. Beide Ansätze zielen aus unterschiedlichen Perspektiven auf das Verstehen der Wechselwirkungen ab. Verschiedene physiologische Bedingungen ergeben diese Wechselwirkungen. Die *Top-down* Strategie analysiert von einer globalen Perspektive aus, durch Zerteilen des Gesamtsystems in seine kleineren interagierenden Teile. Die *Bottom-up* Strategie hingegen spezifiziert zunächst die individuellen Komponenten und Wechselwirkungen im Detail. Anschließend werden die Einzelteile zu einem großen System zusammengefasst, um basierend auf den bekannten Einzelteilen, die biologischen Zusammenhänge zu untersuchen.

In der vorliegenden Arbeit findet die *Bottom-up* Strategie Anwendung, um den Metabolismus eines biologischen Systems zu analysieren. Zunächst wird ein zu untersuchender Organismus ausgewählt und die zu lösende Frage formuliert. Die Rekonstruktion erfordert die folgenden organismen-spezifischen Informationen, die anhand eines kleinen metabolischen Modells in Abbildung 2.5 dargestellt werden.



**Abbildung 2.5:** Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Import-  $R_0$ , einer Exportreaktion  $R_4$ , zwei irreversiblen  $R_1, R_3$  und einer reversiblen Reaktion  $R_2$  (Doppelpfeil, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist), sowie drei Metaboliten  $A, B, C$ . Die entsprechende stöchiometrische Matrix  $N$  (unten links) und die Gleichungen der dynamischen Massenbilanz (unten rechts) sind dargestellt.

Zunächst werden die Reaktionen, die Stoffwechselwege bilden und zugehörige Enzyme und Transportproteine ( $R_0, R_1, R_2, R_3, R_4$ ) identifiziert. Die Substrate und Produkte der Reaktionen ( $A, B, C$ ) werden zusammen mit den stöchiometrischen Koeffizienten definiert. Reaktion  $R_3$  nutzt das Substrat  $C$ , um 2 Moleküle Produkt  $B$  zu bilden. Anschließend wird die Thermodynamik der Reaktionen (reversibel, irreversibel) erfasst. Die Reaktionen  $R_1$  und  $R_3$  sind irreversibel (Kanten führen in eine Richtung) und  $R_2$  ist reversibel (Kanten führen in beide Richtungen und bilden somit eine Hin- und Rückreaktion ab). Die subzelluläre Lokalisation der Reaktionen werden hinzugefügt (hier das Kompartiment Zytosol). Intrazelluläre Transportprozesse werden hinzugefügt, um den Metabolitaustausch zwischen unterschiedlichen Kompartimenten zu ermöglichen. Abschließend wird das System durch extrazelluläre Transportprozesse von der Umgebung abgegrenzt (Import- und Export von Metaboliten).  $R_0$  ist eine Importreaktion, die den Metabolit  $A$  aus der Umgebung in das Zytosol importiert und  $R_4$  ist eine Exportreaktion, die den Metabolit  $B$  aus dem Zytosol in die Umgebung abgibt.

Verschiedene Datenbanken unterstützen den Prozess der Rekonstruktion. Eine Übersicht zu Datenbanken, die u. a. Stoffwechselwege beschreiben, kann der Online Ressource *Pathguide* [7] entnommen werden. Detaillierte Informationen zu enzymkatalysierten Reaktionen sind in der Datenbank *BRENDA* *brenda* aufbereitet. Informationen zu Transportprozessen können der Datenbank *TCDB* [134] entnommen werden.

Im Folgenden werden grundlegende Begriffe aus der metabolischen Modellierung erklärt, die im weiteren Verlauf wiederholt Verwendung finden.

**Stöchiometrische Matrix** Ein metabolisches Modell mit  $n$  Reaktionen und  $m$  Metaboliten wird mathematisch in Form einer  $m \times n$  *stöchiometrischen Matrix*  $N$  repräsentiert. Jede Zeile steht für einen von  $m$  Metaboliten und jede Spalte für eine von  $n$  Reaktionen. Die Einträge, ausgehend von den Spalten, entsprechen den stöchiometrischen Koeffizienten der Metabolite, die an einer Reaktion teilnehmen. Die Substrate einer Reaktion besitzen positive Koeffizienten, Produkte negative. Im Fall reversibler Reaktionen werden die Koeffizienten der Hauptrichtung (vom Substrat zum Produkt) in die stöchiometrische Matrix eingetragen. Nimmt ein Metabolit nicht an einer Reaktion teil, wird eine Null geschrieben. Externe Metabolite werden nicht in die Matrix eingetragen [116]. Wenn nicht explizit eine Stöchiometrie angegeben ist, wird ein stöchiometrischer Koeffizient von eins angenommen, siehe Abbildung 2.5 (unten links).

**Dynamische Massenbilanz** Zur Beschreibung der Dynamik können metabolische Modelle in Form dynamischer Massenbilanzgleichungen dargestellt werden. Diese Massenbilanz besagt, dass die Konzentrationsveränderung von Metaboliten über die Zeit exakt der Differenz zwischen der Summe von Reaktionsraten zur Bildung und der Summe des Verbrauchs des Metaboliten entspricht [44]. Die Konzentration eines Metaboliten wird als Variable  $[S_i]$  beschrieben, die sich über die Zeit  $\frac{d[S_i]}{dt}$  ändert. Die Änderung ist abhängig von den Reaktionsraten  $v_i$ , siehe Abbildung 2.5 (unten rechts).

**Massenwirkungsgesetz** Es besagt, dass die Reaktionsrate proportional zur Wahrscheinlichkeit der Kollision der jeweiligen Substrate ist und diese ist wiederum proportional zur Konzentration von Substraten hoch deren stöchiometrischen Koeffizienten [66]. Die Rate einer Reaktion  $S_1 + S_2 \leftrightarrow 2P$  wird dann wie folgt berechnet.

$$v = v_+ - v_- = k_+ \cdot [S_1] \cdot [S_2] - k_- [P^2]$$

Hierbei ist  $v$  die Nettoreaktionsrate,  $v_+$  und  $v_-$  jeweils die Reaktionsrate der Hin- und Rückreaktion und  $k_+$  und  $k_-$  die entsprechenden kinetischen Ratenkonstanten. Die stöchiometrischen Koeffizienten der Substrate sind jeweils 1 und die des Produktes 2. Im chemischen Gleichgewicht gilt  $k_{eq} = \frac{k_+}{k_-} = \frac{\prod P_{eq}}{\prod S_{eq}}$  unabhängig von den Anfangsbedingungen. Die Gleichgewichtskonstante (engl. *equilibrium constant*)  $k_{eq}$  beschreibt das Verhältnis der Substrat- und Produktkonzentrationen im chemischen Gleichgewicht ( $S_{eq}, P_{eq}$ ), einem Zustand mit gleichen Hin- und Rückreaktionsraten.

**Fließgleichgewicht** Im *Fließgleichgewicht* (engl. *steady-state*) ist die Summe der Zu- und Abflüsse des Modells vom Betrag her gleich, sodass keine Änderungen der Konzentrationsvariablen von Metaboliten auftreten. Somit bleiben die Konzentrationen aller Metabolite im Modell über die Zeit konstant. Es wird angenommen, dass das *Fließgleichgewicht* für jedes metabolische Modell, unter konstanten Umgebungsbedingungen, existiert. Für die meisten biologischen Systeme in der Natur ist bekannt, dass diese nahe dem *Fließgleichgewicht* agieren [19].

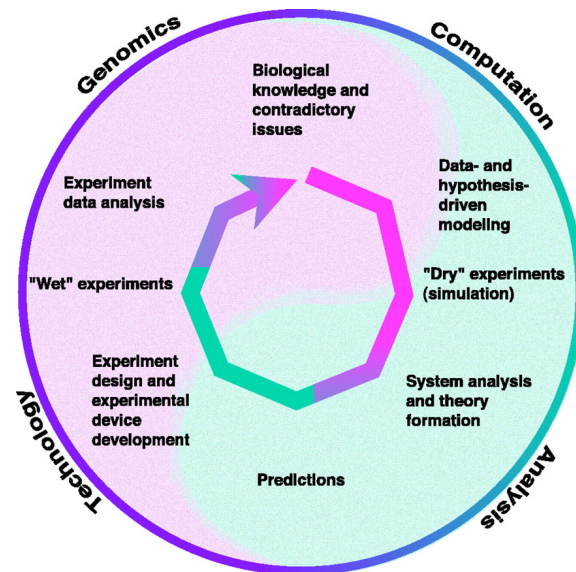
### 2.3.1.3 Erweiterung metabolischer Modelle zur Analyse

Die beschriebene Rekonstruktion bis zu dieser Stelle entspricht einem *topologischen Modell*, welches in der *topologischen Modellierung* auf topologische Eigenschaften untersucht werden kann (siehe Kapitel 3 Abschnitt 3.5). Unter der zusätzlichen Annahme des *Fließgleichgewichtes* wird die Rekonstruktion zum *stöchiometrischen Modell*. Basierend auf der Stöchiometrie der biochemischen Reaktionen und dieser Annahme, können in der *stöchiometrischen Modellierung* strukturelle und optimierungsbasierte Analysen (bedarf zusätzlicher Bedingungen) durchgeführt werden (siehe Kapitel 3 Abschnitt 3.4).

Die Erreichbarkeit und das dynamische Verhalten des Modells können durch die Integration weiterer quantitativer Informationen, der Mengen oder Konzentrationen von Metaboliten und einer möglichen Reaktionsrate in Form von *Petri-Netz Analysen* untersucht werden. Die *Petri-Netz Modellierung* untersucht metabolische Modelle qualitativ, um z. B. Invarianten zu bestimmen und quantitativ, um die Modelldynamik zu untersuchen (siehe Kapitel 3 Abschnitt 3.3).

Werden Reaktionen des metabolischen Modells detailliert durch Reaktionsmechanismen beschrieben, entsteht ein kinetisches Modell. In der *kinetischen Modellierung* beinhalten die Reaktionsmechanismen die Angabe von Metabolitkonzentrationen, Reaktionskinetiken und assoziierten kinetischen Parametern. Im einfachsten Fall wird das *Massenwirkungsgesetz* (engl. *law of mass action*) angenommen, sodass die nötigen Quantitäten Metabolitkonzentrationen und kinetischen Ratenkonstanten sind. Die Simulation eines kinetischen Modells gibt den zeitlichen Ablauf biochemischer Reaktionen an und ermöglicht somit die Analyse der Modelldynamik (siehe Kapitel 3 Abschnitt 3.2).

Die Modellierung beschreibt einen iterativen Prozess. Beginnend mit einer Hypothese wird ein Modell erstellt, um mittels Simulationen Vorhersagen zu treffen. Die Vorhersagen werden experimentell analysiert und schaffen entweder neues Wissen oder bilden eine widersprüchliche Aussage, die zur Anpassung des Modells führt, um erneut Vorhersagen zu treffen. Die Anpassung des Modells erfordert die Verfeinerung des Modells durch die Integration neuer Informationen. Dieser iterative Prozess ist in Abbildung 2.6 dargestellt.



**Abbildung 2.6:** Zyklus der Systembiologie, entnommen aus [85]. Beginnend mit einer Hypothese wird ein Modell erstellt, um mittels Simulationen Vorhersagen zu treffen. Die Vorhersagen werden experimentell analysiert und schaffen entweder neues Wissen oder bilden eine widersprüchliche Aussage, die zur Anpassung des Modells führt, um erneut Vorhersagen zu treffen.

### 2.3.2 Standardisierte Austauschformate für metabolische Modelle

Rekonstruierte metabolische Modelle sollten zur Diskussion mit Wissenschaftlern und zur Verwendung in verschiedenen Computerprogrammen, in standardisierten Formaten austauschbar sein. Im Bereich der Systembiologie und speziell der metabolischen Modellierung werden verschiedene standardisierte Austauschformate verwendet. Aufgrund ihrer Eigenschaften genügen die folgenden Formate den Ansprüchen der metabolischen Modellierung. Die bekanntesten und häufigsten Austauschformate sind *SBML* [74] (engl. *Systems Biology Markup Language*), *CellML* [52] (engl. *Cell Markup Language*) und *BioPax* [37] (engl. *Biological Pathway Exchange*). Diese drei repräsentieren XML-basierte Dateiformate zum Austausch und zur Speicherung von biologischen Modellen. Der *SBML* Standard wird in folgendem Abschnitt erklärt, denn dieser Standard ist am weitesten verbreitet und wird somit als Hauptaustauschformat metabolischer Modelle in dieser Arbeit verwendet. *BioPax* ist speziell für Stoffwechselwege definiert. *CellML* ist *SBML* ähnlich und ermöglicht durch den allgemeinen und flexiblen Syntax die Definition von nicht nur biologischen Modellen.

Ein Austauschformat speziell für Petri-Netze ist *PNML* [69] (engl. *Petri-Net Markup Language*). *PNML* ist wie die oben genannten Austauschformate auch ein XML-basiertes Dateiformat. Auch die Definition von *PNML* ist erweiterbar und nicht auf biologische Petri-Netze beschränkt. Der generische Syntax erlaubt anwendungsspezifische Erweiterungen.

Ist die Visualisierung eines metabolischen Modells von Interesse, wird *SBGNML* [158] (engl. *Systems Biology Graphical Notation Markup Language*) zur Standardisierung des Austausches von Visualisierungen in *SBGN* [100] verwendet. *SBGNML* ist ein XML-basiertes Dateiformat und beinhaltet rein strukturelle Informationen (keine Parameter).

Austauschformate die sowohl die Struktur und das Layout, also auch die gesamten Metainformationen metabolischer Modelle speichern sind *GraphML* [16] (engl. *Graph*

*Markup Language*) und *GML* [70] (engl. *Graph Modelling Language*). *GraphML* ist ein XML-basiertes Dateiformat und *GML* (der Vorgänger von *GraphML*) ist ein ASCII-basiertes Dateiformat. Beide Formate sind durch ihre Erweiterungsmechanismen für anwendungsspezifische Daten sehr flexibel, werden aber in der Systembiologie wenig zum Austausch metabolischer Modelle verwendet.

**SBML** *SBML* ist der erste formale Standard zum Austausch und der Repräsentation mathematischer Modelle biologischer Prozesse. Die weite Unterstützung von *SBML* drückt sich in der Verwendung von ca. 250 Computerprogrammen aus [135]. Es gibt verschiedene *SBML* Level und Versionen, die in *SBML* Spezifikationen beschrieben sind und die Weiterentwicklung dieses Standards dokumentieren [136]. Verschiedene Modelle können durch das prozesszentrierte Konzept unterstützt werden (z. B. kinetische und stöchiometrische Modelle).

In *SBML* werden Modellinformationen in drei Hauptkomponenten eingeteilt, Kompartimente (engl. *compartment*), Metabolite (engl. *species*) und Reaktionen (engl. *reaction*). Die Kompartimente entsprechen begrenzten Reaktionsräumen, in denen Reaktionen ablaufen und Metabolite lokalisiert sind. Die Metabolite beinhalten eine Menge von Substanzen, die an Reaktionen teilnehmen können und in einem Kompartiment lokalisiert sind. Reaktionen sind die Prozesse und können z. B. Transformations-, Transport- oder Bindungsprozesse abbilden und dabei die Menge der Metabolite beeinflussen. Reaktionen werden durch Attribute (z. B. Reversibilität) zusätzlich beschrieben. Außerdem können Reaktionen weitere Elemente beinhalten, z. B. Reaktionsraten aus beliebigen Formeln und Parameter für kinetische Modelle. Den Reaktionen sind Metabolite zugeordnet, die Substrate, Produkte oder Modifikatoren von Reaktionen sind. Metabolite werden ebenfalls durch Attribute näher beschrieben (Lokalisation, Mengenangabe). Generell sind einige Attribute erforderlich und andere optional, siehe *SBML* Spezifikation [136].

Ein hilfreiches Attribut ist der *sboTerm*. *SBO* [26] (engl. *Systems Biology Ontology*) ist eine Sammlung von definierten Bezeichnungen für Entitäten, Prozesse und Ereignisse in systembiologischen Modellen. Ein *sboTerm* definiert *SBML* Elemente eindeutig. Diese Annotation erlaubt *SBML* Elemente mit äquivalenten Konzepten anderer Standards (*SBGN*, siehe Abschnitt 2.3.3.1) in Beziehung zu setzen [18].

### 2.3.3 Visualisierung und Interaktionstechniken

Menschen sind visuelle Wesen und nutzen Zeichen als Schlüssel zum Verständnis. Bevor die geschriebene Sprache formalisiert wurde, gab es bereits die Kommunikation über Bilder (Höhlenmalerei). Visualisierungen dienen der Darstellung und Exploration von Daten. Es gibt die statische und dynamische Visualisierung. Die statische Visualisierung ist eine arrangierte Sicht auf die Daten. Die dynamische, interaktive Visualisierung befähigt zur Exploration von Daten [164]. Zunächst wird eine geeignete Visualisierung für metabolische Modelle beschrieben und anschließend wird auf Interaktionstechniken zur Exploration eingegangen.

#### 2.3.3.1 Visualisierung

In der Visualisierung werden zwei Aspekte unterschieden: Das Modellieren der Daten und das Finden einer geeigneten grafischen Repräsentation [139]. Die Modellierung metabolischer Modelle erfolgt mittels Graphen (siehe Abschnitt 2.2.4). Eine geeignete

grafische Repräsentation metabolischer Modelle ist *SBGN* [100] (engl. *Systems Biology Graphical Notation*).

*SBGN* ist ein Standard zur graphischen Beschreibung biochemischer Prozesse und besteht aus drei Teilsprachen. Die Teilsprache *SBGN-PD* (engl. *Process description*) wird zur Darstellung zeitlicher Abläufe biochemischer Wechselwirkungen in z. B. metabolischen Modellen verwendet. *SBGN-ER* (engl. *Entity relationship*) stellt Beziehungen zwischen Entitäten unabhängig von zeitlichen Verläufen in z. B. Signaltransduktionsmodellen dar. Die dritte Teilsprache *SBGN-AF* (engl. *Activity flow*) beschreibt den Informationsfluß zwischen biochemischen Entitäten in z. B. Genregulationsmodellen. Jede dieser Teilsprachen stellt eine überschaubare Menge von Symbolen bereit. Die in dieser Arbeit verwendete *SBGN-PD* Teilsprache beinhaltet Symbole die der Spezifikation entnommen werden können [112].

Neben der grafischen Repräsentation metabolischer Modelle werden in verschiedenen Analysen unterschiedliche Ergebnisse berechnet (siehe Kapitel 3). Diese Daten können zum einen unabhängig und zum anderen integriert im metabolischen Modell visualisiert werden. Eine Modell unabhängige Visualisierung der Analyseergebnisse, z. B. Metabolitkonzentrationen über die Zeit, können tabellarisch oder zusammengefasst in ein Diagramm dargestellt werden. Es gibt Diagramme verschiedenen Typs, z. B. Balken-, Torten- und Liniendiagramme.

Die Integration der Analyseergebnisse in die Komponenten des Modells (Knoten oder Kanten des Graphen) erlaubt eine direkte Interpretation der Ergebnisse im Kontext des Modells. Hier können zwei Formen der Integration unterschieden werden. Zum einen können die grafischen Attribute wie Farbe, Größe oder Form von Graphenelementen entsprechend der Analyseergebnisse geändert werden. Zur Erhaltung dieser Attribute können zum anderen zusätzliche Elemente wie Text oder Diagramme den Graphenelementen hinzugefügt werden. Beispiele der Visualisierung von Analyseergebnissen sind in Kapitel 5, Abschnitt 5.3.2 in den Abbildungen dargestellt.

### 2.3.3.2 Interaktionstechniken

Das Gebiet der visuellen Analyse (engl. *visual analytics*) wird nach Thomas und Cook [156] als die Wissenschaft der analytischen Argumentation, die durch Interaktionstechniken erleichtert wird, beschrieben. Nach dem Review von Landesberger und Koautoren zur visuellen Analyse großer Graphen [161] gibt es verschiedenen Interaktionstechniken, die in unterschiedliche Kategorien eingeteilt sind. In dieser Publikation werden Interaktionen der Ansicht (engl. *View Interaction*), der visuellen Abstraktion (engl. *Visual Abstraction Interaction*) und der Daten (engl. *Data Interaction*) unterschieden. Die in dieser Arbeit verwendeten Interaktionstechniken *Brushing & Linking* [46, 111] und *Animation* werden der visuellen Abstraktionsinteraktion zugeordnet. Für Informationen zu den genannten Kategorien und zugehörigen Interaktionstechniken sei auf die entsprechende Literatur [147, 156, 164] verwiesen.

**Brushing & Linking** Die Interaktionstechnik *Brushing & Linking* ist eine Standard-Interaktionstechnik zur explorativen Visualisierung [161]. Es wird eine Datenmenge in verschiedenen Perspektiven angezeigt. Das *Linking* beschreibt dabei die Verknüpfung ausgewählter Teile einer Datenmenge in zwei oder mehr Ansichten. Die Selektion und visuelle Hervorhebung der ausgewählten Datenmenge wird durch das *Brushing* realisiert. Eine Selektion von Teilen einer Datenmenge erfolgt durch die Interaktion des Anwenders mit Hilfe der Maus oder Tastatur. Es führt zur automatischen Übertragung der



ausgewählten Daten in allen Ansichten. Das Hervorheben in den verschiedenen Ansichten lenkt die Aufmerksamkeit des Anwenders auf bestimmte Aspekte der Daten. Es gibt verschiedene Möglichkeiten des Hervorhebens wie Farb-, Form- und Größenänderung.

**Animation** Die Interaktionstechnik *Animation* erlaubt den Verlauf der Änderung von Daten z.B. zeitabhängig darzustellen. Ausgehend von einer Ansicht werden die Zwischenschritte bis zur Entstehung der letzten Ansicht dargestellt, wobei die Illusion einer Bewegung entsteht [48]. Die Zwischenschritte führen dabei zum besseren Verständnis der entstehenden Ansicht.



## 3 Modellierungsformalismen metabolischer Modelle

Die Verhaltens- und Funktionsweisen biologischer Systeme zu analysieren ist aufgrund von vielfältigen biochemischen Interaktionen schwierig. Ansätze zur Modellierung und Analyse ermöglichen die komplexen Mechanismen auf Systemebene zu verstehen [73]. Abhängig von der zu lösenden Fragestellung, dem biochemischen Wissen und der Verfügbarkeit experimenteller Daten, können verschiedene Modellierungsformalismen auf Modelle angewendet werden. Geeignete Modellierungsformalismen zur Analyse metabolischer Modelle werden im Abschnitt Stand der Technik evaluiert. Die folgenden Abschnitte stellen die gewählten Modellierungsformalismen, anhand des in den Grundlagen eingeführten metabolischen Beispiel-Modells, vor. Im letzten Abschnitt werden die Formalismen tabellarisch verglichen, siehe Abbildung 3.1.

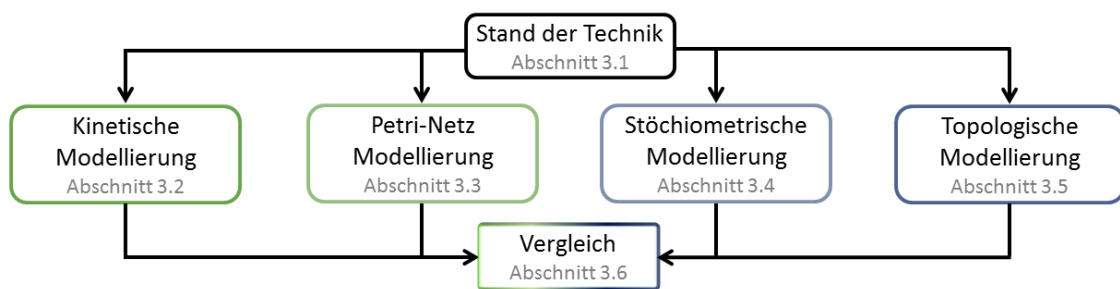


Abbildung 3.1: Struktur des Kapitels Modellierungsformalismen.

### 3.1 Stand der Technik

In der Literatur werden die folgenden etablierten Modellierungsformalismen zur Analyse metabolischer Modelle beschrieben. Der Review von Hübner und Koautoren [73] beinhaltet die Aufarbeitung von 400 Artikeln der Systembiologie. Darin wurden die Modellierungsformalismen zur kinetischen, stöchiometrischen, stochastischen, logischen und Petri-Netz Modellierung als am häufigsten genutzt, bewertet. Die zur Modellierung wichtigen Formalismen sind nach Dandekar und Koautoren [33] die topologische, stöchiometrische und kinetische Modellierung. Machado und Koautoren [106] befassen sich mit Modellierungsformalismen zur Analyse von Signaltransduktions-, Genregulations- und metabolischen Modellen. Darin werden die logische, stöchiometrische, Bayes'sche, Petri-Netz, kinetische, topologische und Agenten-basierte Modellierung, sowie die Verwendung von Zustandsautomaten und zellulären Automaten zur Modellierung in der Systembiologie herausgestellt. Von Steuer und Junker [153] wird die Hierarchie metabolischer Modelle beschrieben und u. a. die topologische, stöchiometrische und kinetische Modellierung präsentiert. Einen weiteren Überblick zu Modellierungsformalismen geben [118, 167] und benennen u. a. topologische, stöchiometrische und kinetische Modellierung.

Für die metabolische Modellierung werden nach Machado und Koautoren [106] hauptsächlich Petri-Netze, stöchiometrische und kinetische Modelle und wenige zelluläre Automaten, sowie Agenten-basierte Modelle verwendet. Aufgrund der in der Literatur beschriebenen, etablierten und häufig verwendeten Modellierungsformalismen wurden zur Abdeckung vielfältiger Analysen in dieser Arbeit die kinetische, Petri-Netz, stöchiometrische und topologische Modellierung ausgewählt. Diese Formalismen erlauben metabolische Modelle unterschiedlicher Größe und enthaltenen Details auf strukturelle und dynamische Eigenschaften zu untersuchen.

## 3.2 Kinetische Modellierung

Basierend auf einem *kinetischen Modell* befasst sich die kinetische Modellierung mit dem zeitlichen Ablauf chemischer Reaktionen.

### 3.2.1 Kinetisches Modell

Ein kinetisches Modell besitzt eine detaillierte und komplexe mathematische Beschreibung der Reaktionen, die ein metabolisches Modell charakterisieren. Jede Reaktion besteht aus einer stöchiometrischen Reaktionsgleichung, deren Reaktionsmechanismus mittels Reaktionskinetiken und assoziierten kinetischen Parametern quantitativ beschrieben wird.

Die Quantitäten eines kinetischen Modells sind *Metabolitkonzentrationen* (z. B. die Konzentration von Metabolit  $[C]$ ) und *Reaktionsraten* (z. B. die Reaktionsrate von Reaktion  $R1$  ist  $v_{R3}$ ) [88]. Die Reaktionsraten können mittels *Reaktionskinetiken*, z. B. der *Michaelis-Menten Kinetik* formuliert werden. Reaktionskinetiken sind abhängig von Metabolitkonzentrationen und kinetischen Parametern, wie Bindungskonstanten und beschreiben den Einfluss von Enzymen und verschiedene Arten der Inhibition. Es können somit komplexe enzymatische Reaktionsmechanismen abgebildet werden. Weiterführende Informationen zur Herleitung von enzymatischen Reaktionsmechanismen und deren Definition sind in der entsprechenden Fachliteratur [11, 15, 88] aufgeführt. Im einfachsten Fall wird basierend auf dem *Massenwirkungsgesetz* die *Massenwirkungskinetik* verwendet (siehe Kapitel 2 Abschnitt 2.3.1.2).

In Form von gewöhnlichen Differentialgleichungen (engl. *ordinary differential equations*, kurz ODEs) beschreibt das kinetische Modell die Konzentrationsdynamik als Veränderung der Konzentration über die Zeit:  $\frac{d[C]}{dt} = -v_{R3}$  und  $\frac{d[B]}{dt} = 2 \cdot v_{R3}$ . Der Abbau von  $C$  erfolgt mit der Rate  $v_{R3}$  und die Bildung von  $B$  erfolgt mit der doppelten Rate. Auch Transportprozesse werden durch kinetische Reaktionsmechanismen beschreiben. Die mathematische Beschreibung eines kinetischen Modells besteht somit aus einem Vektor  $S = (A, B, C)^T$  mit Metabolitkonzentrationen, einem Vektor  $v = (v_{R0}, v_{R1}, v_{R2}, v_{R3}, v_{R4})^T$  mit Reaktionsraten (u. U. komplexe Reaktionskinetiken), einem Parametervektor  $p = (k_0, k_1, k_2, k_3, k_4)^T$  und der *stöchiometrischen Matrix*  $N$ :  $\frac{dS}{dt} = N \cdot v$  (siehe Kapitel 2 Abschnitt 2.3.1.2).

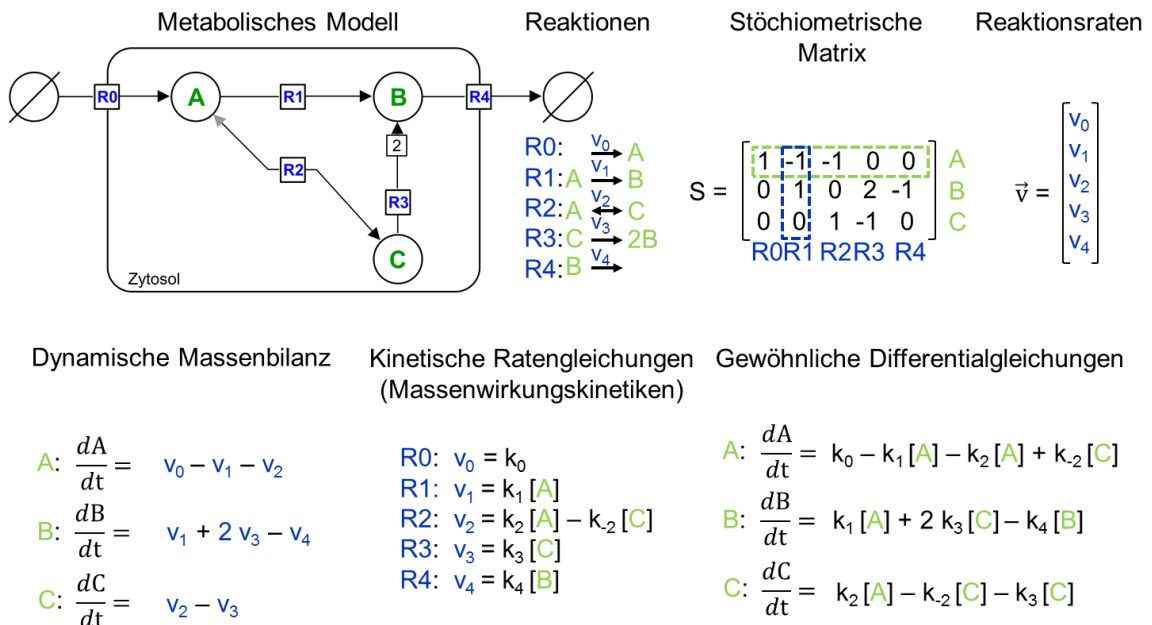
### 3.2.2 Analysen

Die ODE-Systeme in der kinetischen Modellierung sind meist *überbestimmt* (engl. *over-determined*), d. h. die Anzahl der Metabolite bzw. Parameter ist größer als die Anzahl der Reaktionen, an denen diese teilnehmen und können somit nicht analytisch gelöst werden. In diesem Fall finden *numerische Verfahren* Anwendung, die eine approximative Lösung

berechnen. Die Verfahren lassen sich in *explizite*-, *implizite*-, *Ein*- und *Mehrschrittverfahren* unterscheiden [154], sodass die verschiedenen Eigenschaften kinetischer Modelle in der Simulation berücksichtigt werden.

Unter Angabe eines definierten Anfangszustands durch initiale Metabolitkonzentrationen, den Startzeitpunkt, die Schrittweite und den Endzeitpunkt kann ein kinetisches Modell simuliert werden [137] (Zeitreihensimulation). Die *ODEs* können mit diesen Angaben schrittweise gelöst werden (von einem Zeitpunkt zum nächsten). Die Konzentrationsveränderungen der Metabolite werden berechnet, die Metabolitkonzentrationen werden neu gesetzt und eine neue Berechnung findet statt. Dieser Prozess wird solange wiederholt bis der Endzeitpunkt erreicht wird.

Die Konzentration der Metabolite und Reaktionsraten können über die Zeit in ein Diagramm aufgetragen werden. Diese Diagramme können abhängig vom Modell und der Parameter verschiedene Kurvenverläufe darstellen, die unterschiedliche Modellverhalten widerspiegeln (Konvergieren bis zum Fließgleichgewicht oder Divergieren bis zum Halten oder Oszillieren). In Abbildung 3.2 ist die Herleitung der *ODEs* dargestellt. Das Ergebnis der Simulation mit dem Euler Verfahren ist in Abbildung 5.9 dargestellt.



**Abbildung 3.2:** Herleitung der *ODEs* zur kinetischen Analyse am Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Import- *R0*, einer Exportreaktion *R4*, zwei irreversiblen *R1*, *R3* und einer reversiblen Reaktion *R2* sowie drei Metaboliten *A*, *B*, *C*. Reversible Reaktionen werden durch Doppelpfeile dargestellt, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist.

Meistens sind Experimentaldaten zu chemischen Reaktionen nicht präzise, einheitlich und vollständig und das gilt insbesondere für die kinetischen Daten [8]. Die veröffentlichten Daten stammen aus unterschiedlichen Organismen oder verschiedenen Genotypen und werden über mehrere Jahre produziert. Außerdem werden Reaktionskinetiken selten veröffentlicht, da wichtige Substrate, Cofaktoren und Effektoren gewöhnlich getrennt untersucht werden. Die Kinetik vieler Reaktionen ist fast völlig unbekannt und Modellannahmen sind sehr spekulativ. Daher werden vereinfachte Ansätze verwendet. Semi-quantitative Ansätze wie die *strukturelle kinetische Modellierung* (kurz *SKM*) liegt zwischen der kinetischen und

stöchiometrischen Modellierung. Weitere Informationen zur *SKM* können [153] entnommen werden.

### 3.3 Petri-Netz Modellierung

Der graphische und mathematische Petri-Netz Formalismus ist geeignet um nebenläufige, asynchrone, verteilte Systeme zu modellieren und zu analysieren [8, 22].

#### 3.3.1 Petri-Netz Modell

Ein Petri-Netz ist ein gerichteter, bipartiter Graph (siehe Kapitel 2 Abschnitt 2.2.4) und besteht aus den folgenden Grundelementen [113]:

Es gibt zwei Knotenmengen, die *Plätze* (engl. *Places*) stellen die Ressourcen eines Systems dar und die *Transitionen* (engl. *Transitions*) sind die Ereignisse, die Veränderungen der Zustände von Ressourcen durchführen. Weitere Grundelemente sind gerichtete *Kanten* (engl. *arcs*), die immer unterschiedliche Arten von Knoten verbinden und ein *Gewicht* besitzen. Plätze, von denen gerichtete Kanten zu einer Transition führen, heißen *Vorbedingung* dieser Transition. Wenn die Kante von einer Transition zu einem Platz verläuft, wird von einer *Nachbedingung* gesprochen. In manchen Fällen besitzen Transitionen keine eingehenden Kanten (*Quelltransition*) oder keine ausgehenden Kanten (*Senkentransition*).

Übertragen auf ein metabolisches Modell entsprechen Plätze Metabolite und die beinhalteten Marken zeigen die Präsenz und Menge der Metabolite an. Transitionen entsprechen biochemischen Reaktionen. Die Kanten beschreiben Relationen zwischen einem Metaboliten und der Reaktion. Ist ein Platz Vorbedingung einer Transition, dann ist der Metabolit das Substrat der Reaktion. Das Kantengewicht entspricht der Stöchiometrie. Die Markierung beschreibt die vorliegende Metabolitmenge.

**Dynamik** Petri-Netze können das Verhalten von Modellen (Modelldynamik) beschreiben. Die Plätze  $P$  beinhalten keine oder eine positive Anzahl von *Marken* (engl. *Token*) und bilden eine Petri-Netz *Markierung*  $M$  (auch Zustand genannt). Eine Transition  $T$  ist *aktiviert* (engl. *enabled*), wenn alle Vorbedingungen die nötige Anzahl Marken beinhalten (mehr oder oder entsprechend dem Gewicht der Kanten). Eine Transition *schaltet* (engl. *firing*) oder führt einen Schritt aus, indem Marken entsprechend der Gewichte der Kanten aus allen Vorbedingungen entfernt werden und allen Nachbedingungen Marken entsprechend der Kantengewichte hinzugefügt werden. Eine aktivierte Transition kann schalten (muss aber nicht). Die Dynamik eines Petri-Netzes beschreibt den Ablauf biochemischer Reaktionen im metabolischen Modell.

**Erweiterungen** Anders als in der kinetischen Modellierung können Petri-Netze, abhängig von der gewählten Petri-Netz Klasse, qualitative und quantitative Modelle repräsentieren. Qualitative Petri-Netze sind Platz/Transitions-Netze (kurz *P/T-Netze*), die außer der Stöchiometrie und einer möglichen Menge Metabolite keine weiteren quantitativen Informationen enthalten. Es existieren verschiedene Petri-Netz Erweiterungen, die deren Ausdrucksstärke erhöhen, aber auch die Analysen erschweren [8]. Einige dieser quantitativen Petri-Netze enthalten Erweiterungen wie *Test-* und *Inhibitorkanten*. Testkanten aktivieren Transitionen und werden nicht durch Transitionen verändert. Inhibitorkanten werden z. B. verwendet um Transitionen zu inhibieren. Enthält der zugehörige Platz eine Marke, wird die Transition inhibiert und kann nicht schalten. Somit wird die inhibitorische Funktion

einiger Substanzen auf eine Reaktion modelliert. Zur Modellvereinfachung und um Schleifen (engl. *self-loops*) zu verhindern, wird die explizite Repräsentation von Enzymen unterlassen. Dieser Ansatz eignet sich bei metabolischen Modellen, denn die Enzymkonzentrationen ändern sich über die Zeit nicht [8]. Enzyme werden alternativ als Parameter von Reaktionen behandelt.

Andere Petri-Netz Erweiterungen bilden funktionale Petri-Netze (*FPNs*) und definieren kontinuierliche Transitionsschaltraten als Funktionen der Markierung verbundener Plätze. Farbige Petri-Netze (*CPNs*) erweitern Marken um Datenwerte (z. B. Farben) und Kanten um Ausdrücke. Zeitgesteuerte Petri-Netze (*TPNs*) und stochastische Petri-Netze (*SPNs*) beeinflussen das Schalten von Transitionen. Aktivierte Transitionen werden in *SPNs*, zum Beispiel unter Verwendung einer exponentialverteilten Zufallsvariablen, zeitverzögert geschaltet. Kontinuierliche Petri-Netze (*KPNs*) und hybride Petri-Netze (*HPNs*) erlauben die Koexistenz von diskreten und kontinuierlichen Plätzen und Transitionen (diskrete Transitionen schalten nach einer bestimmten Verzögerung, während aktivierte kontinuierliche Transitionen mit einer kontinuierlichen Rate schalten).

In dieser Arbeit werden erweiterte *P/T-Netze* (ePN) und erweiterte *SPNs* (eSPN) verwendet. Die Erweiterung bezieht sich auf die Möglichkeit, die folgenden Parameter und Kantentypen zu nutzen. Es können kontinuierliche Marken (zum Modellieren von Metabolitkonzentrationen), kontinuierliche Kantengewichte (zum Modellieren von Stöchiometrien aus reellen Zahlen), kontinuierliche Kapazitäten (zum Modellieren von limitierenden Metabolitkonzentrationen) und Inhibitorkanten (zur Modellierung von Inhibitionen) verwendet werden.

Ein 6-Tupel  $ePN = (P, T, K, g, f, m_0)$  wird erweitertes Platz/Transitions-Netz genannt falls [151]:

- $P = \{p_1, p_2, \dots, p_n\}$  ist eine endliche, nicht-leere und disjunkte Menge von Plätzen,
- $T = \{t_1, t_2, \dots, t_m\}$  ist eine endliche, nicht-leere Menge von Transitionen,
- $K = K_{standard} \cup K_{inhibitor}$  ist eine Menge von gerichteten Kanten, mit:
  - $K_{standard} \subseteq (P \times T) \cup (T \times P)$  definiert die Menge der Standardkanten,
  - $K_{inhibitor} \subseteq (P \times T)$  definiert die Menge der Inhibitorkanten
- $g : K \rightarrow \mathbb{R}_{>0}$  ist eine Gewichtsfunktion die jeder Kante eine reelle Zahl größer als 0 (Stöchiometrie) zuordnet,
- $f : P \rightarrow \mathbb{R}_{>0} \cup \{\infty\}$  ist eine Kapazitätsfunktion die jedem Platz eine reelle Zahl größer als 0 (Kapazität) bzw.  $\{\infty\}$  (unbeschränkt) zuordnet,
- $m_0 : P \rightarrow \mathbb{R}_{\geq 0}$ , mit  $m_0(p) \leq f(p)$  ist eine Funktion die jedem Platz eine nicht-negative reelle Zahl (Anfangsmarkierung) zuordnet. Zudem muss diese Markierung kleiner oder gleich der Kapazität sein.

Ein erweitertes stochastisches Petri-Netz (eSPN) geht aus dem ePN hervor, wobei jeder Transition eine stochastische Schaltrate hinzugefügt wird. Bevor eine aktivierte Transition schaltet, verstreicht eine gewisse zufällige Zeitdauer. Diese Zeitdauer ist durch eine exponentialverteilte Zufallsvariable definiert. Jeder Transition ist daher eine stochastische Schaltrate oder äquivalent eine mittlere Schaltdauer zugeordnet. Kehrwerte von Schaltdauern sind Schaltraten, welche Parameter einer Exponentialverteilung sind [110]. Diese stochastische Schaltrate kann durch die Angabe eines zusätzlichen Gewichtes (z. B. der

Reaktionsrate) beeinflusst werden und die stochastische Simulation gerichteter durchführen zu können.

Ein 7-Tupel  $eSPN = (P, T, K, G, F, M_0, R)$  wird erweitertes, stochastisches Petri-Netz genannt falls [110]:

- $P, T, K, g, f, m_0$  Definition wie in ePN,
- $R : T \rightarrow \mathbb{R}_{\geq 0}$ , ist eine Funktion, die jeder Transition eine stochastische Schaltrate zuordnet.

**Eigenschaften** Petri-Netz Eigenschaften werden zum einen durch die Struktur (*markierungsunabhängig*) und zum anderen durch die Markierung (*markierungsabhängig*) festgelegt. Die folgenden Ausführungen sind dem Buch [151] entnommen. Markierungsabhängige Eigenschaften sind z. B. *Erreichbarkeit*, *Beschränktheit*, *Lebendigkeit*. Eine Markierung eines Petri-Netzes heißt *erreichbar*, falls es eine Schaltabfolge der Transitionen gibt, welche die Anfangsmarkierung in diese Markierung überführt. Ein Petri-Netz heißt *beschränkt*, wenn es für jeden Platz eine obere Schranke  $k$  gibt, sodass in allen Zuständen nie mehr als  $k$  Marken auf einem Platz liegen. Ein Petri-Netz heißt *lebendig*, wenn alle Transitionen lebendig sind, d. h. jede Transition ist in jeder erreichbaren Markierung aktivierbar. Metabolische Modelle in Form von Petri-Netzen sind oftmals *unbeschränkt* (Aufnahme und Abgabe von Metaboliten über die Systemgrenze), lebendig (alle Reaktionen können ablaufen) und die Folgezustände sind erreichbar.

Markierungsunabhängige Eigenschaften sind z. B. *Invarianten*, *Konflikte*, *Fallen* und *Verklemmungen*. Nach Starke [151] wird als Invariante eines Systems eine Eigenschaft bezeichnet, die unabhängig von der konkreten Schaltabfolge erhalten bleibt. Aus der Existenz einer Invariante können Rückschlüsse auf andere Eigenschaften gezogen werden. Man unterscheidet zwischen *Platzinvarianten* (kurz P-Invarianten) und *Transitionsinvarianten* (kurz T-Invarianten). P-Invarianten geben eine Wichtung für eine in jeder erreichbaren Markierung konstante Markierungssumme ab. Die P-Invarianten können als Erhaltungsregeln eines metabolischen Modells interpretiert werden, denn diese veranschaulichen die Substanzerhaltung von Metaboliten. Eine T-Invariante beschreibt eine gültige Schaltabfolge, bei der der Anfangszustand wieder erreicht wird. Im Kontext metabolischer Modelle entsprechen T-Invarianten den Reaktionen von Hauptstoffwechselwegen, deren Ausführung zum Gleichgewichtszustand führen. Invarianten stellen somit die Beziehung zwischen der Struktur des Petri-Netzes und dem Verhalten des Systems dar.

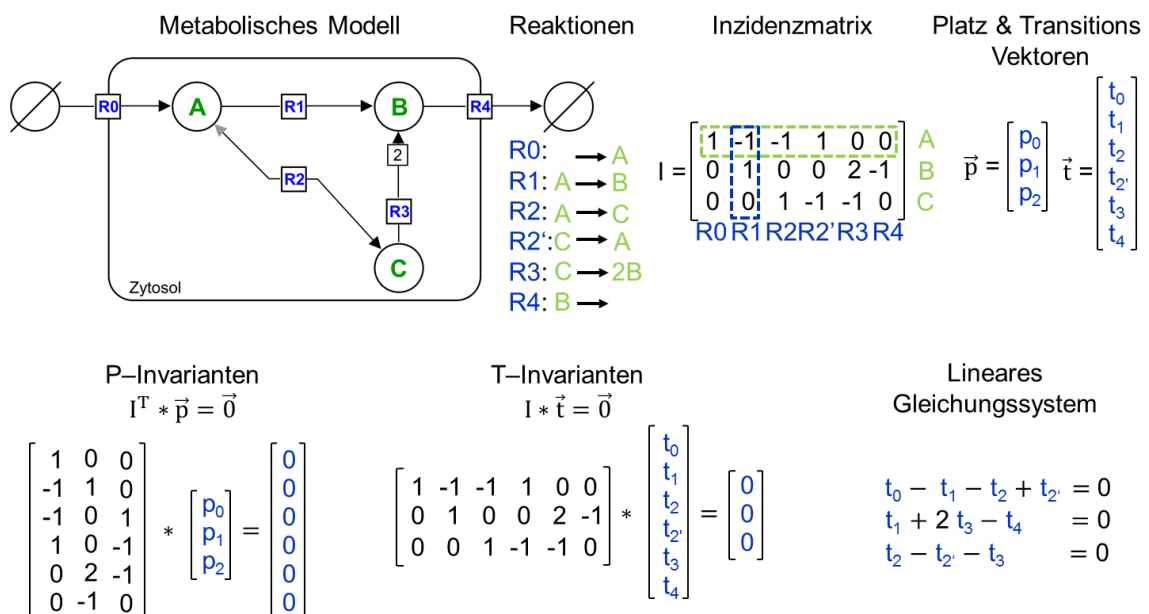
Ein Konflikt besteht, wenn mindestens zwei Transitionen existieren, die sich einen Vorplatz teilen. Übertragen auf ein metabolisches Modell beschreibt ein Konflikt die Dilatation in der mindestens zwei Reaktionen ein und denselben Metaboliten verbrauchen. Eine Verklemmung entsteht, wenn alle Transitionen die Marken einer nicht-leeren Menge von Plätzen hinzufügen und auch Marken aus dieser Menge entnehmen (u. U. mehr entnehmen als hinzufügen). Wird in einem metabolischen Modell ein Speicher von Metaboliten beschrieben, wird dieser bei einer Verklemmung entleert. Im Gegensatz zur Verklemmung entsteht eine Falle, wenn alle Transitionen, die Marken einer nicht-leeren Menge von Plätzen entnehmen, auch Marken dieser Menge hinzufügen (u. U. mehr hinzufügen als entnehmen). Bei einer Falle wird in einem metabolischen Modell der Speicher von Metaboliten angelegt.



### 3.3.2 Analysen

Die Bestimmung von Petri-Netz Eigenschaften kann mittels *qualitativen* oder *quantitativen* Analysen (Simulation) erfolgen.

**Qualitative Analysen** Die qualitativen Analysen von Petri-Netzen können in *statische* und *dynamische* Analysemethoden unterteilt werden. Eine statische Analysemethode ist z. B. die *Invariantenanalyse*. Die Lösung eines linearen Gleichungssystems ermöglicht die Berechnung von Invarianten durch z. B. das *Gauß'sche Eliminationsverfahren*. Zur Berechnung der P-Invarianten ist die Formel  $I^T \times \vec{p} = \vec{0}$  und zur Berechnung der T-Invarianten ist es  $I \times \vec{t} = \vec{0}$ .  $I$  ist die Inzidenzmatrix (entspricht der stöchiometrischen Matrix, siehe Kapitel 2 Abschnitt 2.3.1.2),  $\vec{p}$  ist der P-Invariantenvektor,  $\vec{t}$  ist der T-Invariantenvektor und  $\vec{0}$  ist ein Vektor bestehend aus Nullen. In Abbildung 3.3 ist die Herleitung der Gleichungen zur Invariantenanalyse des metabolischen Modell Beispiels dargestellt. Eine der T-Invarianten ist als Ergebnis in Abbildung 5.10 gezeigt.



**Abbildung 3.3:** Herleitung der Gleichungen zur Invariantenanalyse am Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Import-  $R_0$ , einer Exportreaktion  $R_4$ , zwei irreversiblen  $R_1, R_3$  und einer reversiblen Reaktion  $R_2$  sowie drei Metaboliten  $A, B, C$ . Reversible Reaktionen werden durch Doppelpfeile dargestellt, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist.

Die Anzahl der zu berechnenden Invarianten wächst exponentiell mit der Größe des Petri-Netzes [89].

Eine dynamische Analysemethode ist z. B. die *Erreichbarkeitsanalyse*. Es wird beginnend mit der Anfangsmarkierung die Menge der in der aktuellen Markierung aktivierten Transitionen ermittelt und jeweils die Folgemarkierung berechnet. Die möglichen Markierungen werden durch Knoten im *Erreichbarkeitsgraphen* dargestellt und der Übergang einer Markierung zu ihrer Folgemarkierung wird als gerichtete Kante im Graphen vermerkt. Für jede Folgemarkierung wird dieser Vorgang wiederholt. Nur ein beschränktes Petri-Netz besitzt einen endlichen Erreichbarkeitsgraphen. Die Erreichbarkeit in einem unbeschränkten Petri-Netz wird durch einen *Überdeckungsgraphen* dargestellt.

Eine *pseudo Endlichkeit*  $\omega$  dient der Markierung von unbeschränkten Plätzen. Unabhängig von der Größe eines Petri-Netzes kann aufgrund willkürlicher Verschachtelung parallel ablaufender Transitionen eine sehr große Erreichbarkeitsmenge entstehen. Die Berechnung der Erreichbarkeit führt zur *Zustandsraum Explosion* (engl. *state space explosion*). In diesem Fall ist die Erreichbarkeit unberechenbar [8].

**Quantitative Analysen** Mit zunehmender Größe metabolischer Modelle wird es schwierig, qualitative Analysen durchzuführen. In diesem Fall werden die Eigenschaften von Petri-Netzen durch quantitative Analysen (Simulationen) untersucht. Im Gegensatz zu *ePNs*, deren Schaltabfolge nicht-deterministisch ist, ermöglichen *eSPNs* ein stochastisches Schaltverhalten. Die Quantitäten sind *stochastische Schaltraten* (möglicherweise gewichtet) und Marken. In der Simulation wird der Übergang von einem Zustand zum folgenden Zustand zeitlich geordnet, durch das zufallsgesteuerte Schalten von Transitionen beschrieben. Im Vergleich zu deterministischen kinetischen Modellen bilden *eSPNs* metabolische Modelle mit einer feineren Granularität ab. Es können interessante individuelle Verhaltensweisen beobachtet werden, die in deterministischen kinetischen Modellen eher den Durchschnittsfall erfassen [8]. Des Weiteren beschreiben biochemische Reaktionen metabolischer Modelle grundsätzlich stochastische Prozesse. Eine Abschätzung der Zeit, wann die nächste Reaktion durchgeführt wird (wie in deterministischen kinetischen Modellen), ist somit nicht möglich [129].

Unter Angabe eines definierten Zustandes durch Anfangsmarkierung und einer Gewichtung der stochastischen Schaltraten (ein Gewicht kann, muss aber nicht angegeben werden) wird ein *eSPN* simuliert. Dazu erhält jede aktivierte Transition eine Zufallszahl, die multipliziert mit dem Gewicht der Transition zu einer Schaltwahrscheinlichkeit verrechnet wird. Die aktivierte Transition mit der höchsten Wahrscheinlichkeit wird schalten und somit wird ein Simulationsschritt im Petri-Netz ausgeführt. Die Markierung des *eSPNs* wird entsprechend angepasst. Anschließend wird die Menge der aktivierten Transitionen erneut berechnet und mit Zufallszahlen versehen, um die Schaltwahrscheinlichkeit zu bestimmen. Dieser Prozess wird solange wiederholt bis die angegebene Anzahl der Simulationsschritte erreicht wird.

Wird die Simulation mit den gleichen Anfangsbedingungen wiederholt, entsteht ein anderes Ergebnis, denn die Zufallszahlen sind verschieden. Die Bildung des Durchschnitts über mehrere Simulationsläufe resultiert in der Annäherung an die präzise Lösung. Die Metabolitkonzentrationen und Reaktionsraten (Schaltraten) können über die Anzahl der Simulationsschritte in Diagrammen aufgetragen werden. Diese Diagramme können abhängig vom Modell und der Parameter verschiedene Kurvenverläufe darstellen, die unterschiedliche Modellverhalten widerspiegeln.

### 3.4 Stöchiometrische Modellierung

Zur kinetischen Modellierung und stochastischen Petri-Netz Analyse werden detaillierte, quantitative Beschreibungen benötigt. Die stöchiometrische Modellierung hingegen basiert auf der Stöchiometrie biochemischer Reaktionen, die den Ausgangspunkt für Analysen bildet.

### 3.4.1 Stöchiometrisches Modell

Ein stöchiometrisches Modell wird mittels stöchiometrischer Reaktionsgleichungen beschrieben. Es werden dynamische Massenbilanzgleichungen (siehe Kapitel 2 Abschnitt 2.3.1.2) formuliert und aufgrund der Annahme eines *Fließgleichgewichtes*, (siehe Kapitel 2 Abschnitt 2.3.1.2) die in dieser Arbeit verwendeten stöchiometrischen Analysen ermöglicht.

### 3.4.2 Analysen

Es werde zwei Kategorien stöchiometrischer Analysen unterschieden, die strukturellen und die optimierungsbasierten Analysen metabolischer Modelle [104].

**Strukturell** Zu den strukturellen Analysen gehören die Bestimmung von z.B. elementaren Flussmoden (kurz *EFM*) [142], *Extreme Pathways* (kurz *EP*) [138] und *Dead-End* Metaboliten [124]. Die ersten beiden Analysen eignen sich z. B. zur Ermittlung nicht redundanter Stoffwechselwege in metabolischen Modellen. Weitere Informationen siehe angegebene Referenzen. Die *Dead-End* Analyse identifiziert Metabolite, die nur verbraucht oder produziert werden (engl. *dead end metabolites*) durch die Untersuchung der Verbindungen von Metaboliten in der stöchiometrischen Matrix (siehe Kapitel 2 Abschnitt 2.3.1.2).

**Optimierungsbasiert** Ein anderer Ansatz *Dead-End* Metabolite und die dazugehörigen blockierten Reaktionen zu identifizieren, ist die optimierungsbasierte *Gap-Finding* Analyse [95]. Das Modell wird hinsichtlich der Möglichkeit, dass jede Reaktion einen Fluss tragen könnte (unter beliebigen Aufnahmebedingungen), untersucht. Auf diese Weise werden blockierte Reaktionen ermittelt, die direkt oder indirekt mit einem oder mehreren *Dead-End* Metaboliten zusammenhängen. Die *Dead-End* und *Gap-Finding* Analysen dienen in metabolischen Modellen der Verifizierung. Besitzt ein Modell blockierte Reaktionen fehlen zusätzliche Reaktionen, die diese Metabolite erzeugen oder produzieren, damit ein voll funktionsfähiges Modell entsteht.

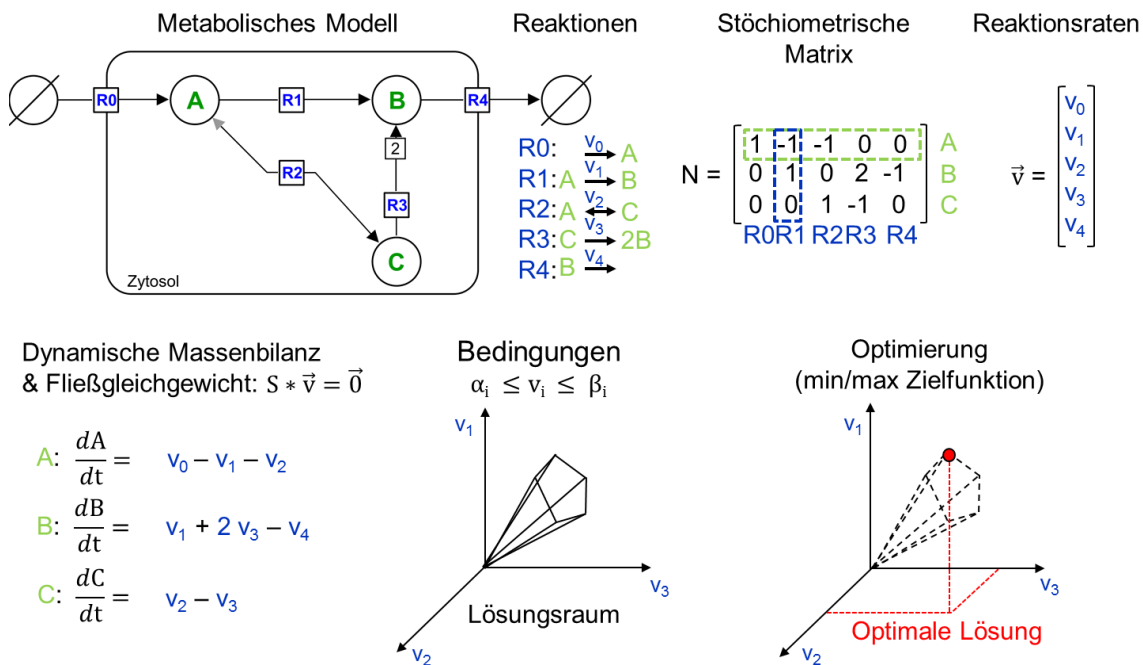
Weitere optimierungsbasierte Analysen sind z. B. die *metabolische Flussanalyse* (kurz *MFA*) [152], *Flussbilanzanalyse* (kurz *FBA*) [115], *Flussvariabilitätsanalyse* (kurz *FVA*) [108], *Robustheitsanalyse* (*RA*) [45] und *Knockout-Analyse* (*KA*) [115]. Die metabolische Flussanalyse ist eine quantitative Analyse, die intrazelluläre Stoffflüsse während des Fließgleichgewichtes (siehe Kapitel 2 Abschnitt 2.3.1.2), abhängig von z. B. 13-C Markierungsdaten, bestimmt. Weitere Informationen siehe [152]. Die optimierungsbasierten Analysen *FBA* [115] und deren verwandte Analysen (*FVA*, *RA*, *KA*) [104] überwinden die Beschränkungen, ungenügend experimentelle Daten zur Verfügung zu haben. Die *FBA* berechnet den Metabolitfluss (Flussverteilung) durch das metabolische Modell, wobei es möglich ist, die Wachstumsrate des modellierten Organismus oder die Rate eines biotechnologisch wichtigen Metaboliten abzuschätzen.

**Bedingungen** Diese Abschätzung (Optimierung) findet unter der Einbeziehung von Bedingungen (engl. *constraints*) statt:

1. Annahme des Fließgleichgewichtes:  $\frac{dS}{dt} = N \cdot v = 0$ ,
2. Thermodynamik ((Ir)reversibilität von Reaktionen),

3. Eingeschränkte Kapazität von Reaktionen:  $lb_i \leq v_i \leq ub_i$  (untere, obere Schranke),
4. Biomassezusammensetzung (Lipide, Proteine, Kohlehydrate) und
5. Externe Bedingungen (Aufnahme-/Abgaberraten bestimmter Metabolite)

Durch die erste Bedingung reduziert sich das aus dem Modell hergeleitete *ODE*-System in ein System linearer algebraischer Gleichungen. Für ein metabolisches Modell ist dieses Gleichungssystem *unterbestimmt*, d. h. die Anzahl der Reaktionen  $n$  ist größer als die Anzahl der teilnehmenden Metabolite  $m$ :  $n > m$  [86]. Der Lösungsraum wäre ohne die weiteren Bedingungen unendlich groß. Die angegebenen Bedingungen begrenzen die möglichen Flussverteilungen auf eine eingeschränkte Menge, erreichen aber keine eindeutige Lösung. Die Bestimmung einer Flussverteilung wird als *LP-Problem* (engl. *linear programming*), formuliert. (Zur Bestimmung der eindeutigen Flussverteilung kann eine zweite Optimierung zur Minimierung des Gesamtflusses hinzugefügt werden, welches dann als *QP-Problem* (engl. *quadratic programming*) formuliert wird.) Die Idee ist eine *Zielfunktion*  $Z$  gemäß der Stöchiometrie und den Bedingungen zu optimieren (maximieren oder minimieren):  $Z = \sum_{i=1}^r c_i^T \cdot v_i \rightarrow \max$  ( $c_i$  ist ein Gewichtsvektor für die individuellen Reaktionen, die zur Zielfunktion beitragen). Beispiele für Zielfunktionen sind Maximierung der ATP Produktion, Minimierung der Nährstoffaufnahme, maximale Ausbeute eines gewünschten Produktes oder eine maximale Wachstumsrate [47, 141]. Das Ergebnis der *FBA* ist eine Flussverteilung  $v$ , die die Zielfunktion optimiert. In Abbildung 3.4 wird die Herleitung zur Berechnung der optimalen Lösung des metabolischen Modell Beispiels dargestellt. Die erzielte Flussverteilung unter Maximierung der Zielfunktion *R4* wird mittels Kantenskalierung in Abbildung 5.11 gezeigt.



**Abbildung 3.4:** Herleitung zur Berechnung der optimalen Lösung mittels *FBA* am Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Import- *R0*, einer Exportreaktion *R4*, zwei irreversiblen *R1*, *R3* und einer reversiblen Reaktion *R2* sowie drei Metaboliten *A*, *B*, *C*. Reversible Reaktionen werden durch Doppelpfeile dargestellt, wobei die Haupttrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist.

Basierend auf *FBA* können auch die Analysen *FVA*, *RA* und *KA* durchgeführt werden. Bei der *FBA* kann es zu einer optimalen Lösung kommen die nicht eindeutig ist, (z. B. durch die Nutzung alternativer Stoffwechselwege). Die *FVA* gibt für jede Reaktion den maximalen und minimalen Flusswert ( $v_i^{min}$ ,  $v_i^{max}$ ) an, der zur Bestimmung der optimalen Flussverteilung möglich ist. Die *RA* ermittelt den optimalen Wert der Zielfunktion mit Hilfe der Variation einer Reaktion. Das Ergebnis ist die Sensitivität der Zielfunktion zu dieser Reaktion. Diese wird häufig als Sensitivitätskurve dargestellt. Im metabolischen Modell kann die *RA* z. B. den Effekt der Nährstoffaufnahme auf die Wachstumsrate ermitteln [45]. Die *KA* wird zur Bestimmung essentieller Reaktionen und deren Einfluss auf die Flussverteilung verwendet. Dazu wird die Kapazität einer Reaktion auf 0 gesetzt, d. h. die untere und obere Schranke der Reaktion sind 0 [115]. Das Ergebnis der *KA* sind Sensitivitätswerte aller Reaktionen auf die *stillgelegte* Reaktion. Weiterführende Informationen können [60] entnommen werden.

### 3.5 Topologische Modellierung

Zum besseren Verständnis der Bedeutung zentraler Elemente, der Struktur und Robustheit gegenüber Störungen, werden metabolische Modelle nach topologischen Eigenschaften analysiert. Aus der topologischen Sicht ist ein metabolisches Modell ein gerichteter bipartiter Graph (siehe Abschnitt 2.2.4 und Abschnitt 2.3.1.2). In der Literatur werden zur topologischen Analyse hauptsächlich uniparte *Metabolit-* und *Reaktionsgraphen* verwendet [153], sodass auch in dieser Arbeit diese beiden Graphen genutzt werden.

#### 3.5.1 Topologisches Modell

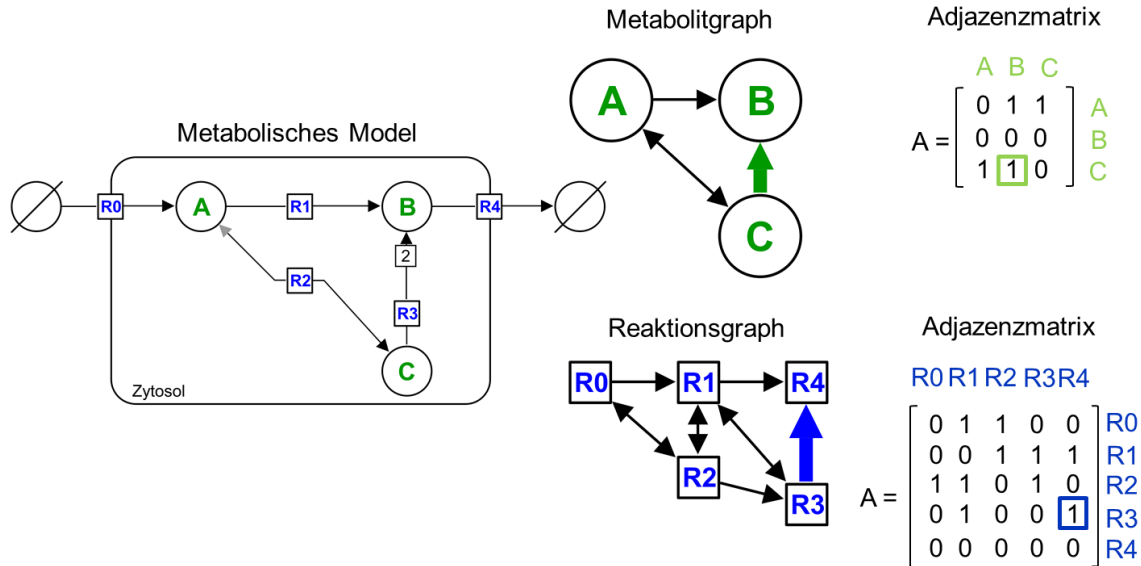
Aus einem metabolischen Modell wird durch Projektion auf je eine der beiden Knotenmengen der zugehörige *Metabolit-* bzw. *Reaktionsgraph* erstellt (siehe Kapitel 2 Abschnitt 2.2.4). Der *Metabolitgraph* besteht aus Metabolitknoten, die über gerichtete Kanten (Reaktionen) miteinander verbunden sind. Eine Kante wird zwischen zwei Metaboliten gezogen, wenn diese im metabolischen Modell mit ein und derselben Reaktion verbunden ist. Ein *Metabolitgraph* beschreibt somit, welcher Metabolit in einen anderen umgewandelt werden kann. Der *Reaktionsgraph* besteht aus Reaktionsknoten, die über gerichtete Kanten (Metabolite) miteinander verbunden sind. Eine Kante wird zwischen zwei Reaktionen gezogen, wenn diese im metabolischen Modell mit ein und demselben Metaboliten verbunden ist. Ein *Reaktionsgraph* beschreibt somit, welche Reaktion Vorgänger oder Nachfolger einer anderen Reaktion ist. In diesen speziellen Graphen wurden nach Literaturangaben [91, 153, 170] bisher keine Kantengewichte (Stöchiometrie) berücksichtigt.

#### 3.5.2 Analysen

Die topologische Modellierung beinhaltet verschiedene Analysen, die auf Basis der *Adjazenzmatrix* (siehe Kapitel 2 Abschnitt 2.2.4) berechnet werden. Die Analyse der *kürzesten Wege* kann nach Angabe eines Ausgangsmetaboliten (-reaktion) und eines Zielmetaboliten (-reaktion) den kürzesten Stoffwechselweg ermitteln.

Die Erstellung einer Rangliste von Graphenelementen (Knoten und Kanten) ermöglicht zentrale Elemente im metabolischen Modell zu identifizieren, die z. B. für die Lebensfähigkeit essentiell sind. Basierend auf verschiedenen Konzepten bestimmt die *Zentralitätsanalyse* eine solche geordnete Liste von Zentralitätswerten. Die häufig verwendeten Zentralitäten zur Analyse metabolischer Modelle beruhen auf Nachbarschaft (engl. *degree*), Wege (engl. *closeness*) und Beobachtbarkeit (engl. *Shortest-Path Betweenness*) [91, 79, 162].

In Abbildung 3.5 wird die Herleitung der zwei Graphen (Metabolit- und Reaktionsgraph) und der entsprechenden Adjazenzmatrizen als Basis für Analysen dargestellt. Das Ergebnis der topologischen Analyse, *Shortest-Path Betweenness* Zentralität des Reaktionsgraphen, wird in Abbildung 5.12 gezeigt.




**Abbildung 3.5:** Herleitung der Metabolit- und Reaktionsgraphen am Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Import-  $R_0$ , einer Exportreaktion  $R_4$ , zwei irreversiblen  $R_1$ ,  $R_3$  und einer reversiblen Reaktion  $R_2$  sowie drei Metaboliten  $A$ ,  $B$ ,  $C$ . Reversible Reaktionen werden durch Doppelpfeile dargestellt, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist. Zur Übersichtlichkeit (Reduktion von Kanten und Überschneidungen) werden in den Metabolit- und Reaktionsgraphen antiparallele Kanten zwischen den Knoten durch Doppelpfeile dargestellt. Markierte Kanten zeigen den Zusammenhang zwischen Metabolit- bzw. Reaktionsgraph und den zugehörigen Eintrag in den Adjazenzmatrizen.

### 3.6 Vergleich der Modellierungsformalismen

Die Verwendung der vorgestellten Modellierungsformalismen ist hauptsächlich von der Verfügbarkeit detaillierten Wissens über Interaktionsmechanismen metabolischer Prozesse abhängig. Entsprechend dieser Experimentaldaten kann aus den verschiedenen Modellierungsformalismen gewählt werden. In Tabelle 3.1 sind die Stärken und Schwächen der vorgestellten Modellierungsformalismen zusammengefasst.

Die kinetische Modellierung beschreibt biochemische Prozesse detailliert, quantitativ und benötigt daher oftmals schwierig zu beschaffende kinetische Ratengleichungen und Parameter. Aufgrund dessen ist die kinetische Modellierung im Allgemeinen auf kleinere Modelle limitiert, führt jedoch zu quantitativen Vorhersagen und spiegelt das dynamische Verhalten des zugrunde liegenden biologischen Systems wider. Infolge der Erweiterbarkeit von Petri-Netzen ist die Petri-Netz Modellierung sehr mächtig und ermöglicht neben der qualitativen auch die quantitativen Analysen. Zum Beispiel werden *eSPNs* zur stochastischen Simulation verwendet, um die Dynamik und quantitative Vorhersagen zu treffen. Allerdings wird die Erreichbarkeitsanalyse als Beispiel einer qualitativen Analyse aufgrund der Beachtung von z. B. Inhibitorkanten komplizierter.

**Tabelle 3.1:** Stärken (grünes Plus) und Schwächen (rotes Minus) der vorgestellten Modellierungsfomalismen. Abhängig von detailliertem Wissen über Interaktionsmechanismen entstehen kleine, eher quantitative Modelle und große, meist qualitative Modelle. Verschiedene Modellierungsfomalismen verwenden unterschiedliche Modelle (Zusammenhang farblich gekennzeichnet) und liefern ergänzende Erkenntnisse über das zugrunde liegende biologische System.



Kinetische Modellierung	Petri-Netz Modellierung	Stöchiometrische Modellierung	Topologische Modellierung
<ul style="list-style-type: none"> <li>+ Detaillierte quantitative Beschreibung</li> <li>+ Quantitative Vorhersagen und dynamisches Verhalten</li> </ul>	<ul style="list-style-type: none"> <li>+ Quantitative Beschreibung</li> <li>+ Erweiterbar (eSPN, ...)</li> <li>+ Quantitative Vorhersagen und dynamisches Verhalten</li> </ul>	<ul style="list-style-type: none"> <li>+ Fließgleichgewichtannahme</li> <li>+ Quantitative Vorhersagen</li> <li>+ Große Modelle</li> </ul>	<ul style="list-style-type: none"> <li>+ Nur topologische Information</li> <li>+ Große Modelle</li> </ul>
<ul style="list-style-type: none"> <li>- Kinetische Ratengleichungen und Parameter benötigt</li> <li>- Kleine Modelle</li> </ul>	<ul style="list-style-type: none"> <li>- Erweiterungen ermöglichen quantitative Vorhersagen</li> <li>- Erweiterungen erschweren qualitative Analysen</li> </ul>	<ul style="list-style-type: none"> <li>- Statische Beschreibung</li> <li>- Keine dynamischen Eigenschaften</li> </ul>	<ul style="list-style-type: none"> <li>- Keine dynamischen Eigenschaften</li> </ul>

Die stöchiometrische Modellierung nutzt die Annahme eines *Fließgleichgewichtes*, damit die statische Beschreibung der biochemischen Prozesse ausreichend ist, quantitative Vorhersagen zu treffen. Somit ist die stöchiometrische Modellierung für große Modelle anwendbar, aber limitiert in Aussagen zur Dynamik. Topologische Informationen sind der Ausgangspunkt für die topologische Modellierung. Uneingeschränkt durch die Größe von Modellen nutzt zum Beispiel die Zentralitätsanalyse verschiedene Konzepte zur Untersuchung zentraler Elemente, basierend auf den Metabolit- oder Reaktionsgraphen. Dynamische Eigenschaften des zugrunde liegenden biologischen Systems werden in der topologischen Modellierung nicht hergeleitet.

Zur Nutzung des Potentials der verschiedenen Modellierungsfomalismen für metabolische Modelle ist eine Integration dieser Formalismen anzustreben. Die Integration ermöglicht die Verwendung verschiedener Analyseergebnisse in anderen Modellierungsfomalismen und limitiert somit die Schwächen der voneinander unabhängigen Formalismen. Auf diese Weise können trotz fehlender detaillierter, quantitativer Beschreibungen dennoch Rückschlüsse auf die Verhaltens- und Funktionsweise zugrunde liegender biologischer Systeme gezogen werden. Im folgenden Kapitel 4 werden zunächst bestehende Ansätze zur metabolischen Modellierung evaluiert.





## 4 Evaluierung bestehender Ansätze zur metabolischen Modellierung

Dieses Kapitel beschäftigt sich mit der Evaluierung bestehender Ansätze zur metabolischen Modellierung, die in verschiedenen Computerprogrammen implementiert sind. Es werden Anwendungen vorgestellt und verglichen, die die am meisten verwendeten Modellierungsformalismen (siehe Kapitel 3), d. h. kinetische, Petri-Netz, stöchiometrische und topologische Modellierung unterstützen. Zunächst werden Kriterien eingeführt, die zur Anwendung der Modellierungsformalismen erforderlich sind. Anschließend erfolgt die Evaluierung von spezifischen und integrierten Anwendungen. Die spezifischen Anwendungen ermöglichen die Verwendung einer der vier vorgestellten Modellierungsformalismen. Der integrierte Ansatz hingegen erlaubt die Kombination dieser vier Formalismen in einem Computerprogramm. Eine Zusammenfassung von erweiterten Anforderungen erläutert am Ende dieses Kapitels die Notwendigkeit der Entwicklung einer Methodik, die die vier vorgestellten Modellierungsformalismen integriert und eine einheitliche Visualisierung metabolischer Modelle vornimmt (siehe Abbildung 4.1).

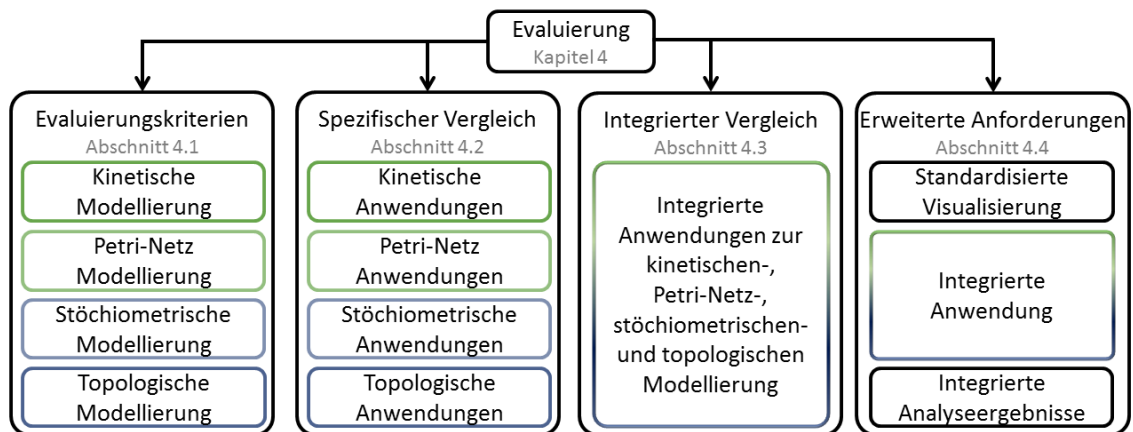


Abbildung 4.1: Struktur des Kapitels Evaluierung.

### 4.1 Evaluierungskriterien

In der Literatur werden nach Machado und Koautoren [106] u. a. die Anforderungen Visualisierung, topologische, qualitative und quantitative Analyse sowie Standardisierung erhoben. Auch in dem Artikel von Dandekar und Koautoren [33] wurden die Kriterien Visualisierung und die Verwendung verschiedener Modellierungsformalismen (z. B. topologisch, stöchiometrisch und kinetisch) zur metabolischen Modellierung als erforderlich herausgestellt. Basierend auf diesen und weiteren Literaturrecherchen zum aktuellen technischen Entwicklungsstand [73, 118, 153, 167] (erläutert in Kapitel 3) und aus Gesprächen mit Biologen und Bioinformatikern zur Aufnahme erforderlicher Kriterien an

Computerprogrammen zur metabolischen Modellierung der benannten Ansätze, wurden die folgenden Kriterien zusammengetragen.

Metabolische Modelle werden hauptsächlich unter der Verwendung von *SBML* [74] (engl. *Systems Biology Markup Language*) zwischen verschiedenen Computerprogrammen ausgetauscht (Erläuterungen zu *SBML* siehe Kapitel 2 Abschnitt 2.3.2). *SBML* ist ein weit verbreiteter Standard für metabolische Modelle [135] und wird somit als Kriterium Hauptaustauschformat festgelegt.

Die metabolischen Modelle sollten intuitiv und einfach interpretierbar visualisiert werden. Ein geeigneter Standard zur Visualisierung ist *SBGN* [100] (engl. *Systems Biology Graphical Notation*) (siehe Kapitel 2 Abschnitt 2.3.3.1). Diese graphische Repräsentation soll die Integration verschiedener Analyseergebnisse der unterschiedlichen Modellierungsfomalismen ermöglichen. Zur Exploration von Analyseergebnissen der Modellierungsfomalismen werden an geeigneten Stellen Interaktionstechniken (siehe Kapitel 2 Abschnitt 2.3.3.2) verwendet.

Die Kriterien an die verschiedenen Modellierungsfomalismen werden im Folgenden erläutert.

### 4.1.1 Kinetische Modellierung

In der kinetischen Modellierung (Erklärungen zu verwendeten Fachbegriffen siehe Kapitel 3 Abschnitt 3.2) werden hauptsächlich Zeitreihensimulationen und das Lösen des Fließgleichgewichtes vorgenommen [3, 125, 130]. Die kinetische Analyse metabolischer Modelle soll unter Verwendung verschiedener numerischer Lösungsverfahren erfolgen und die Einstellung des Fließgleichgewichtes überprüfen. Eine geeignete Parametrisierung, wie a) die Änderung von initialen Modellparametern, b) Angaben wie Start- und Endzeitpunkt und c) die Schrittweite der Simulation, sind notwendig. Die Analyseergebnisse sollen in Form von Diagrammen in die graphische Repräsentation integriert werden, tabellarisch exportierbar sein und in einem Überblicksdiagramm zusammengefasst werden.

### 4.1.2 Petri-Netz Modellierung

Metabolische Modelle in Form von Petri-Netzen werden zur qualitativen und quantitativen Analyse herangezogen [8, 22]. Die meist verwendeten Petri-Netz Klassen sind P/T-Netze zur qualitativen Analyse und kontinuierliche Petri-Netze (KPNS), hybride Petri-Netze (HPNS) und stochastische Petri-Netze (SPNS) zur quantitativen Analyse [8] (Erklärungen zu verwendeten Fachbegriffen siehe Kapitel 3 Abschnitt 3.3). Petri-Netze werden in *PNML* [69] (engl. *Petri-Net Markup Language*) definiert, sodass ein Kriterium die Verwendung von *PNML* ist. Die zu verwendenden Petri-Netz Klassen sind ePNs und eSPNs, mit der Möglichkeit, Inhibitorikanten und kontinuierliche Quantitäten zu verwenden (Definition siehe Kapitel 3 Abschnitt 3.3). Die qualitative Petri-Netz Analyse soll die Invarianten- und Erreichbarkeitsanalyse ermöglichen. Die entstehenden Platz- und Transitionsinvarianten, als auch der Erreichbarkeits- oder Überdeckungsgraph, sollen mittels Interaktionstechniken mit dem metabolischen Modell verbunden werden. Zur Analyse der Modelldynamik sollen Simulationen durchgeführt werden, deren Ergebnisse in Form von Diagrammen in die graphische Repräsentation integriert werden und tabellarisch exportierbar sind.

### 4.1.3 Stöchiometrische Modellierung

Die stöchiometrische Modellierung gliedert sich in strukturelle Analysen (z. B. Analyse elementarer Flussmoden [142], *Dead-End* Analyse [124]) und optimierungsbasierte Analysen (z. B. metabolische Flussanalyse [152], Flussbilanzanalyse [115]) metabolischer Modelle [104] (Erklärungen zu verwendeten Fachbegriffen siehe Kapitel 3 Abschnitt 3.4). Die erforderlichen Analysen sind optimierungsbasierte Methoden wie die Flussbilanzanalyse *FBA* und verwandte Methoden (Flussvariabilitätsanalyse *FVA* [108], Robustheitsanalyse *RA* [45] und Knockout-Analyse *KA* [115]), die nach Lakshmann und Koautoren [99] eine der am häufigsten verwendeten Methoden für metabolische Modelle ist. Zur Validierung metabolischer Modelle werden die *Dead-End* und die *Gap-Finding* Analysen gefordert. Resultierende Analyseergebnisse sollen in die graphische Repräsentation des metabolischen Modells integriert werden und mittels Interaktionstechniken explorierbar sein.

### 4.1.4 Topologische Modellierung

Die topologische Modellierung ermöglicht zentrale Elemente, die Struktur und Robustheit gegenüber Störungen metabolischer Modelle aufzuklären. Diese Analysen erfolgen hauptsächlich anhand des *Metabolit-* oder *Reaktionsgraphen* [153] (Erklärungen zu verwendeten Fachbegriffen siehe Kapitel 3 Abschnitt 3.5). Zur Umwandlung eines metabolischen Modells in diese Graphen sollten geeignete Graphumwandlungen bereitgestellt werden. Die Analyse dieser Graphen soll Auskunft über die Modellstatistik und die kürzesten Wege zwischen ausgewählten Knoten geben. Weiterhin soll die Anwendung verschiedener Zentralitätsmaße möglich sein. Entstehende Analyseergebnisse sollen in Tabellen zusammengefasst werden und mit einer Interaktionstechnik mit dem metabolischen Modell verbunden werden.

Die festgelegten allgemeinen und speziellen Kriterien resultieren in der Auswahl der im Folgenden untersuchten quelloffenen, eigenständigen (engl. stand-alone) und nicht kommerziellen Computerprogramme. Zunächst werden die Anwendungen evaluiert, die spezialisiert für einen Modellierungsformalismus entwickelt wurden.

## 4.2 Vergleich spezifischer Modellierungsansätze

Zahlreiche quelloffene, eigenständige und nicht kommerzielle Computerprogramme zur metabolischen Modellierung sind derzeit verfügbar. Die meisten sind zur Unterstützung eines spezifischen Modellierungsformalismus entwickelt wurden. Obwohl innerhalb der Anwendungen für entsprechende Formalismen zum Teil überlappende Funktionalitäten bestehen, führt jede dieser Anwendungen zur Lösung spezieller Probleme.

### 4.2.1 Kinetische Modellierung

In der kinetischen Modellierung stehen *CellDesigner*, *Cellware*, *Dizzy*, *SBMLSimulator* und *COPASI* im Vergleich, siehe Tabelle 4.1.

Die Anwendungen *BIOCHAM*, *iBioSim*, *BioSPICE*, *BioUML*, *SBW* und *VANTED* sind ebenfalls Computerprogramme zur kinetischen Modellierung, die aufgrund der Unterstützung weiterer Modellierungsformalisten im Vergleich der integrierten Anwendungen in Abschnitt 4.3 erläutert werden.

**Tabelle 4.1:** Vergleich kinetischer Anwendungen.

Anwendung		CellDesigner	Cellware	Dizzy	SBMLSimulator	COPASI
Austauschformate						
SBML	Import	x	x	x	x	x
	Export	x	x	x	x	x
Visualisierung						
SBGN		x				
Graphvisualisierung		x	x			
Interaktiv		x				x
Graphintegration						
Kinetische Modellierung						
Simulation	Determ.	x	x	x	x	x
	Stoch.	x	x	x		x
Parametrisierung		x	x	x		

*CellDesigner* [50] ist ein Modellierungsprogramm zur Visualisierung, Erstellung und Simulation biochemischer Modelle. Es verwendet die Standards *SBML* als Modellaustauschformat und *SBGN-PD* zur graphischen Repräsentation von Modellen (wobei diese Version von *SBGN-PD* eine Weiterentwicklung ist und nicht dem allgemeinen Standard entspricht). Vor allem überzeugt *CellDesigner* mit der Möglichkeit ein Modell graphisch zu definieren. Mit der Integration von *SBML ODE Solver (SOSlib)*, *Copasi* und *SBW (Jarnac)* können Modelle simuliert werden. Die Interaktionstechnik Animation wird verwendet, um durch Auswahl von verschiedenen Parametern Schieberegler zu generieren. Deren Verschiebung zwischen zwei Werten (Min und Max) ermöglicht eine Anpassung der Simulationsergebnisse (Kurven im Diagramm).

Die Anwendung *Cellware* [38] bietet wie *CellDesigner* die Möglichkeit biochemische Modelle graphisch zu definieren, visualisieren und simulieren. Ebenso können *SBML* Modelle importiert und exportiert werden. Neben den deterministischen, stochastischen und hybriden Algorithmen zur Simulation können zusätzlich Parameterschätzungen durchgeführt werden. Des Weiteren kann die Homologie von Stoffwechselwegen überprüft werden. Analyseergebnisse werden in einem Report zusammengefasst, der sowohl Tabellen als auch Diagramme beinhaltet.

*Dizzy* [123] wird als Simulationsprogramm für kinetische Modelle beschrieben und bietet eine große Auswahl verschiedener deterministischer und stochastischer Lösungsverfahren. Mehrere Parameter ermöglichen eine vielfältige Anpassung für die Simulation. Zur Bestimmung der Parameter mit dem größten Einfluss auf die Kinetik des Modells kann eine Sensitivitätsanalyse durchgeführt werden. *Dizzy* ermöglicht *SBML* Modelle auszutauschen, verfügt aber über keine eigene Visualisierung. Modelle können aber in *Cytoscape* visualisiert werden.

Der *SBMLSimulator* [83] wird als effizienter und leicht verständlicher Simulator für *SBML* Modelle beschrieben. Dieser interpretiert ein Modell als ein ODE System, welches mit neun verschiedenen Solvern, durch numerische Integration, simuliert werden kann. Außerdem gibt es die Möglichkeit Parameter zu schätzen.

*COPASI* [71] ist ein Simulator für biochemische Modelle der zudem auch eine Vielzahl von Analysen unterstützt (z. B. Steady-state Analyse und Parameterschätzungen). Modelle können in *SBML* geladen und exportiert werden und die Simulation kann deterministisch und stochastisch erfolgen. Wie *CellDesigner* verfügt *COPASI* ebenfalls über eine Animation

der Simulationsergebnisse, in Form einer Graphvisualisierung, dessen Knotenfarbe oder -größe sich entsprechend der Veränderung der Metabolitkonzentration anpasst. Diese Visualisierung ist aber nur möglich, wenn es Layoutinformationen im *SBML* gibt. *COPASI* ist sehr gut dokumentiert und kompensiert somit die nicht sehr intuitive Bedienung.

#### 4.2.2 Petri-Netz Modellierung

Die Petri-Netz Programme *PIPE2*, *WoPed*, *Renew*, *Jarp*, *QPME*, *Peabrain* und *Snoopy* werden verglichen, siehe Tabelle 4.2.

**Tabelle 4.2:** Vergleich Petri-Netz Anwendungen.

Anwendung		PIPE2	WoPeD	Renew	Jarp	QPME	Peabrain	Snoopy
Austauschformate								
SBML	Import							x
	Export							x
PNML	Import	x	x	x	x	x	x	
	Export	x	x	x	x	x	x	x
Visualisierung								
SBGN								
Graphvisualisierung		x	x	x	x	x	x	x
Interaktiv		x	x	x	x			
Graphintegration								
Petri-Netz Modellierung								
Klasse	P/T	x <sub>e</sub>	x	x	x	x	x	x <sub>e</sub>
	SPN	x				x	x	x
Analyse	Inv.	x			x			
	Err.	x	x	x				
Simulation		x	x	x	x	x	x	x

Des Weiteren verfügen *BIOCHAM*, *iBioSim*, *SBW*, *Cytoscape* und *VANTED* ebenfalls über Petri-Netz Simulations- oder Analysemethoden, die in der Betrachtung der integrierten Anwendung (siehe Abschnitt 4.3) untersucht werden.

*PIPE2* [40] ist ein Computerprogramm zur Erstellung, Analyse und Simulation von Petri-Netzen ((erweiterte)-P/T-Netze, (erweiterte) SPNs). Es unterstützt das Petri-Netz Austauschformat *PNML* und ist durch seinen modularen Aufbau durch die Integration von Modulen einfach erweiterbar. Es besitzt einen graphischen Editor und ist intuitiv zu bedienen. Die Visualisierung der Analyseergebnisse erfolgt tabellarisch, die Simulation wird animiert wiedergegeben.

*WoPed* [43] wird hauptsächlich zur Bearbeitung, Verwaltung, Simulation und Analyse von Geschäftsprozessmodellen (Workflow-Netzen) verwendet. Es ermöglicht aber auch P/T-Netze zu nutzen. *WoPeD* besitzt einen grafischen Editor, eine animierte Simulation ermöglicht die Erreichbarkeit zu analysieren und unterstützt den Austausch von Petri-Netzen durch *PNML*.

*Renew* [49] ist ein graphischer Editor und Simulator, der eine flexible Modellierung auf Basis von Referenznetzen, P/T-Netzen und SPNs ermöglicht.

*Jarp* erlaubt die graphische Erstellung von P/T-Netzen, die zur Berechnung von Invarianten und für die animierte Simulation verwendet werden können.

Für sogenannte Warteschlangen-Petri-Netze stellt *QPME* [92] eine spezielle Modellierungsumgebung bereit. Diese Petri-Netze sind eine Kombination aus Warteschlangen-Netzwerken und SPNs. *QPME* unterstützt auch P/T-Netze und SPNs. Für biologische Anwendungen wurde *QPME* noch nicht verwendet. *QPME* ist aber zur stochastischen Modellierung und Analyse fähig. Es besteht aus zwei Modulen, dem *QPE* (graphischer Editor) und dem *SimQPN* (effizienter Simulator).

Das Programm *Peabrain* [127] basiert auf *PIPE2*. Es ist eine Sammlung von Modulen für Leistungsschätzung und Ressourcenoptimierung in SPNs. Die stochastische Simulation von Petri-Netzen hat *PIPE2* der Entwicklung von *Peabrain* zu verdanken.

*Snoopy* [68] ist eine Anwendung zur Erstellung und Simulation von diversen Petri-Netzen (z. B. (erweiterte) P/T-Netze, (erweiterte) SPNs, TPNs, CPNs, HPNs), die außerdem ineinander umgewandelt werden können. Ein Editor ermöglicht die Konstruktion von Petri-Netzen. Unter den hier verglichenen Petri-Netz Programmen ist *Snoopy* das Einzige, welches u. a. *SBML* und *PNML* (nur Export) unterstützt. Es ermöglicht u. a. deterministische als auch stochastische Simulationen.

### 4.2.3 Stöchiometrische Modellierung

In der stöchiometrischen Modellierung stehen *GEMSiRV*, *Omix*, *YANA*, *SurreyFBA*, *SBRT* und *BioOpt* im Vergleich, siehe Tabelle 4.3.

Außerdem werden die Anwendungen *OptFlux*, *BioSPICE*, *BioUML* und *VANTED*, die auch stöchiometrische Analysen ermöglichen, im integrierten Vergleich in Abschnitt 4.3 erläutert.

*GEMSiRV* [103] ist ein Programm zur Rekonstruktion, Simulation und Visualisierung metabolischer Modelle. Es unterstützt die Rekonstruktion metabolischer Modelle durch die Möglichkeit eine Referenzdatenbank zu integrieren. Die *Dead-end* und *Gap-Finding* Analysen unterstützen die Modellverfeinerung. Die meist verwendeten optimierungsbasierten Methoden *FBA*, *RA*, *KA* und *FVA* können angewendet werden. Zur Visualisierung werden die erhobenen Resultate in das Modell integriert, z. B. werden die Flusswerte an die Reaktionknoten geschrieben und die Richtung des Flusses durch zusätzliche Pfeile visualisiert.

*Omix* [41] ermöglicht hochwertige Modellvisualisierungen und ist mit umfangreichen Datenvisualisierungsfunktionen ausgestattet. Mit Hilfe von *FBA*- und *EFM*-Plugins können Modelle analysiert und die Ergebnisse interaktiv exploriert werden. Zur Visualisierung der *EFM* wird die *Brushing & Linking* Interaktionstechnik verwendet. Flusswerte werden mittels Kantenskalierungen dargestellt. Weiterhin gibt es die Möglichkeit, zeitabhängige Daten zu animieren, die aus anderen Programmen stammen.

Die *YANA* [144] Computerprogramme erlauben Modelle tabellarisch zu definieren und diese anschließend zu visualisieren. Modelle können mit *FBA*, *RA* und *EFM* analysiert werden und die Analyseergebnisse werden mittels Kantenskalierungen in das Modell integriert. Zusätzlich ist es möglich, experimentelle Datensätze wie Genexpressions- oder Enzymaktivitätsdaten, in der *EFM* zu verwenden.

*SurreyFBA* [54] wird zur Analyse und Visualisierung metabolischer Modelle verwendet. Modelle werden tabellarisch erstellt oder bearbeitet und anschließend visualisiert. Im Vergleich zu den untersuchten Anwendung beinhaltet *SurreyFBA* die meisten

**Tabelle 4.3:** Vergleich stöchiometrischer Anwendungen.

Anwendung		GEMSiRV	Omix	YANA	SurreyFBA	SBRT	BioOpt
Erweiterung			EFM-, FBA- Plugin				SBML2BioOpt
Austauschformate							
SBML	Import	x	x	x	x	x	x
	Export	x	x	x			
Visualisierung							
SBGN							
Graphvisualisierung		x	x	x	x		
Interaktiv			x		x		
Graphintegration		x	x	x	x		
Stöchiometrische Modellierung							
Dead-ends		x			x		
Gap-Finding		x			x		
FBA		x	x	x	x	x	x
RA		x		x	x		
KA		x			x	x	x
FVA		x			x	x	x
Met. Engineering							
EFM			x	x	x	x	x

Analyseverfahren. Die generierten Analyseergebnisse werden als Diagramme oder numerische Werte in die Knoten des Modells integriert.

*SBRT* [168] ist die Abkürzung für *Systems Biology Research Tool*. Die Anwendung stellt verschiedenste Methoden zur Analyse von Modellen, wie *FBA*, *FVA*, *KA* und *EFM* bereit. *SBRT* ist eine Kommandozeilen-Anwendung, die eine Vielzahl von verschiedenen Dateien benötigt, um Modelle analysieren zu können. Die Visualisierung eines Modells und der Analyseergebnisse ist nicht möglich.

*BioOpt* [31] ist eine Kommandozeilen-Anwendung zur Analyse (*FBA*, *KA*, *FVA* und *EFM*) von metabolischen Modellen. Es gibt keine Möglichkeit der Rekonstruktion, Bearbeitung und Visualisierung von Modellen. Die erstellten Analyseergebnisse werden in Form einer Textdatei exportiert.

#### 4.2.4 Topologische Modellierung

Die Programme zur topologischen Modellierung *BiologicalNetworks*, *Gephi*, *Visone* und *CentiBiN* werden hier verglichen, siehe Tabelle 4.4.

Des Weiteren verfügen *Cellware*, *OptFlux*, *Cytoscape* und *VANTED* ebenfalls über topologische Analysemethoden, die in der Betrachtung der integrierten Anwendung (siehe Abschnitt 4.3), untersucht werden.

*BiologicalNetworks* [93] ist neben *CentiBiN* ein Netzwerkanalyseprogramm für biologische Netzwerke. Es verarbeitet als einziges Programm in diesem Vergleich *SBML* Dateien. Ein

**Tabelle 4.4:** Vergleich topologischer Anwendungen.

Anwendung		Biological Networks	Gephi	Visone	CentiBiN
Austauschformate					
SBML	Import	x			
	Export	x			
Visualisierung					
SBGN					
Graphvisualisierung		x	x	x	x
Interaktiv			x	x	x
Graphintegration		x	x	x	x
Topologische Modellierung					
Graphumwandlung					
Zentralitäten			x	x	x
Kürzester Weg		x	x		
Statistik		x	x	x	x

graphischer Editor bietet die Möglichkeit, biologische Netzwerke zu erstellen und zu bearbeiten. In der Netzwerkanalyse wird die Netzwerkstatistik erhoben und die kürzesten Wege analysiert. Weiterhin ist die Verwendung von experimentellen Daten zur Exploration im Netzwerkkontext (z. B. Genexpressionsdaten) möglich.

*Gephi* [9] ist ein Netzwerkanalyseprogramm und überzeugt durch vielfältige dynamische Netzwerkvisualisierungen sehr großer Netzwerke (20.000 Knoten). Neben der Analyse der Netzwerkstatistik erlaubt *Gephi* Knotenzentralitätsanalysen (*BetweennessCentrality*, *ClosenessCentrality*, *Eccentricity*, *PageRank*, *HITS*) und die Berechnung kürzester Wege zwischen zwei Knoten.

Die Anwendung *Visone* [17] wird zur Analyse und Visualisierung sozialer Netzwerke verwendet. *Visone* besitzt eine interaktive grafische Benutzeroberfläche und auf soziale Netzwerke zugeschnittene Netzwerkvisualisierungen. Die Analysen basieren auf Netzwerkstatistik und 12 verschiedenen Kantenzentralitäten. Analyseergebnisse werden direkt in das Netzwerk integriert.

*CentiBiN* [78] ist ein Programm für die Berechnung und Exploration von Zentralitäten in biologischen Netzen. Es berechnet 17 verschiedene Zentralitäten für gerichtete oder ungerichtete Netzwerke von lokalen (direkte Nachbarschaft), als auch globalen Messungen. *CentiBiN* visualisiert die Verteilung zentraler Elemente im Netzwerk mit Hilfe der *Brushing & Linking* Interaktionstechnik. Es ist der Vorgänger des *VANTED* Addons *CentiLib* [64].

**Zusammenfassung** Die Anwendungen, welche spezialisiert für einen Modellierungsformalismus entwickelt wurden, beinhalten ein breites Spektrum von Optionen für die spezielle Analyse. Zum Beispiel beinhaltet ein Programm zur kinetischen Analyse, wie *Copasi*, eine Vielzahl veränderlicher Parameter, Lösungsverfahren und Analysemethoden. Jedes dieser Programme ist für diesen einen Formalismus ausgereift. Soll ein metabolisches Modell mit verschiedenen Modellierungsformalisen untersucht werden, wird die Installation unterschiedlicher Anwendungen notwendig. Der Nutzer muss dafür die Verwendung der verschiedenen Programme erlernen und sich an neue Menüführungen gewöhnen. Manche Anwendungen visualisieren das zu analysierende Modell auf unterschiedliche Weise und andere sehen ganz davon ab. Eine integrierte Anwendung,



die modellzentriert ist und mehr als nur einen dieser Modellierungsformalisten unterstützt, würde vielfältigere Analysen metabolischer Modelle in einem Computerprogramm ermöglichen. Diese Anwendungen werden im folgenden Abschnitt verglichen.

### 4.3 Vergleich integrierter Modellierungsansätze

Neben den Anwendungen, die speziell für einen bestimmten Modellierungsformalismus entwickelt wurden, gibt es immer mehr Computerprogramme, die verschiedene Formalismen beinhalten. In Tabelle 4.5 werden acht dieser Anwendungen verglichen.

Alle der aufgeführten Computerprogramme unterstützen unter anderem den Import und Export von metabolischen Modellen definiert in *SBML* (*BioSpice* nur den Import). Sie erlauben mit Hilfe eines graphischen Editors Modelle zu erstellen und zu bearbeiten.

Zunächst werden die Anwendungen *BIOCHAM* und *iBioSim* eingeführt, die kinetische als auch Petri-Netz Analysen ermöglichen. *BIOCHAM* [20] unterstützt mit dem integrierten graphischen Editor die Entwicklung, Erweiterung, Korrektur, Reduzierung und Verknüpfung metabolischer Modelle, visualisiert in *SBGN*. Auf der Petri-Netz Ebene werden P/T-Netze unterstützt, die auf P-Invarianten untersucht werden können. Die kinetische Analyse umfasst deterministische und stochastische Analysen, die weitreichend parametrisiert werden können. Des Weiteren erlaubt *BIOCHAM* Parametersuche, *Modelchecking* und statistische Analysen.

*iBioSim* [114] ist ein Computerprogramm zur Modellierung, Analyse und zum Entwurf genetischer Schaltungen (engl. genetic circuits). Es können auch metabolische Modelle verwendet werden. Die Analysen beinhalten die Simulation von P/T-Netzen sowie die deterministische und stochastische kinetische Analyse mit geeigneter Parametrisierung. Beiden Anwendungen *BIOCHAM* und *iBioSim* fehlen Interaktionstechniken zur Exploration der Analyseergebnisse.

Die Kombination von stöchiometrischer- und topologischer Modellierung erfüllt das Programm *OptFlux* [126] mit der Aktivierung seiner Erweiterungen *Efm4OptFlux*, *OptKnock4OptFlux*, *Celldesigner4OptFlux* und *TNA4OptFlux*. Metabolische Modelle werden importiert, bearbeitet und exportiert. *OptFlux* ermöglicht stöchiometrische Analysemethoden, wie *FBA*, *KA*, *FVA*, *Met.Engineering* und *EFM* anzuwenden. Weiterhin werden topologische Analysemethoden, wie z. B. die Berechnung der kürzesten Wege und Zentralitäten (*HITSHubs*, *HITSAuths*, *Closeness*) angeboten und die Modellstatistik erhoben.

*BioSpice* und *BioUML* unterstützen kinetische- und stöchiometrische Modellierung. Ein graphischer Editor ermöglicht in beiden Anwendungen die Konstruktion von Arbeitsabläufen zur Analyse von Modellen. *BioSpice* [94] benötigt die Erweiterungen *Jarnac*, *Fluxor* und *MOMA* und kann somit die stöchiometrischen Analysemethoden *FBA* und *Met.Engineering* sowie deterministische und stochastische Simulationen durchführen. Die Anwendung *BioUML* [90] visualisiert zudem die Modelle in *SBGN* und unterstützt das *SGBNML* Austauschformat. Die Analysemethoden sind in *BioUML* vielfältiger und können besser parametrisiert werden.

Aufgrund der Erweiterbarkeit durch Plugins stellen die *System Biology Workbench (SBW)*, *Cytoscape* und *VANTED* bereits drei der vier Modellierungsformalisten bereit. *SBW* [12] umfasst den gesamten Modellierungsablauf von der Rekonstruktion über Bearbeitung, Simulation, Analyse und Visualisierung von Modellen durch die Aktivierung von Plugins. Die beiden Plugins *PKN2e* und *ISBLibrary* erlauben den Austausch von Modellen im Petri-Netz Austauschformat *PNML* und ermöglichen quantitative Petri-Netz Analysen für

**Tabelle 4.5:** Vergleich integrierter Anwendungen. Fußnoten:  $e$  erweiterte Petri-Netze,  $p$  nur Platzinvarianten,  $q$  quadratische Optimierung.

Anwendung	BIOCHAM	iBioSim	OptFlux	BioSPICE	BioUML	SBW	Cytoscape	VANTED
Erweiterung			EFM-, OptKnock-, CellDesigner-, TNA4OptFlux	Jarnac, Fluxor, MOMA		RoadRunner, FBASBW, PKN2e, ISBLibrary, Jarnac	CentiLib, CentiScaPe, BiNoM, Fern, CySBGN, CySBML	CentiLib, FBASimVis, PetriNet, SBGN-ED
Austauschformate								
SBML	Import	x	x	x	x	x	x	x
	Export	x	x	x	x	x	x	x
PNML	Import					x		x
	Export					x		x
SBGNML	Import				x		x	x
	Export				x		x	x
Visualisierung								
SBGN	x				x		x	x
Graphvisualisierung	x	x	x	x	x	x	x	x
Interaktiv						x	x	x
Graphintegration								x
Kinetische Modellierung								
Simulation	Determ.	x	x		x	x	x	
	Stoch.	x	x		x	x	x	
Parametrisierung	x	x			x			
Petri-Netz Modellierung								
Klasse	P/T	x	x				$x_e$	x
	SPN						x	x
Analyse	Inv.	$x_p$						x
	Err.							x
Simulation		x				x	x	x
Stöchiometrische Modellierung								
Dead-ends								
Gap-Finding								
FBA			x	x	x	x		$x_q$
RA								$x_q$
KA			x					$x_q$
FVA			x					x
Met. Engineering			x	x				
EFM			x			x		
Topologische Modellierung								
Graphumwandlung			x					
Zentralitäten			x				x	x
Kürzester Weg			x				x	x
Statistik			x				x	x

erweiterte P/T-Netze und erweiterte SPNs. Die berechneten Analyseergebnisse werden in Diagrammen zusammengefasst und können auch für jeden Zeitschritt im Petri-Netz betrachtet werden (engl. Tokenflow). Ein weiteres Plugin ist *FBASBW*, welches Modelle in *SBGN* visualisiert und die stöchiometrischen Analysen *FBA* und *EFM* unterstützt. Die Analyseergebnisse werden durch Kantenskalierung im Modell repräsentiert. Zur kinetischen Analyse (deterministische und stochastische Simulationen) in *SBW* dienen die Plugins *Jarnac* und *RoadRunner*. Diagramme visualisieren die berechneten Ergebnisse. Des Weiteren bietet *SBW* die Möglichkeit, verschiedene Matrizen des zu untersuchenden Modells direkt zu analysieren und noch weitere kinetische Analysen durchzuführen.

Leider sind die verschiedenen Plugins nicht direkt miteinander verbunden, deshalb stellt *SBW* kein integriertes Computerprogramm dar. Es existiert z.B. kein Hauptanwendungsfenster. Das Wechseln von einem Plugin zum anderen Plugin generiert neue Fenster mit einer anderen Menüführung. Bereits importierte Modelle des einen Plugins werden nicht automatisch in das neu geöffnete Plugin übertragen. Ein fließendes Vorgehen wird daher nicht gewährleistet.

Im Gegensatz zu *SBW* sind *Cytoscape* [145] und *VANTED* integrierte Anwendungen zur Visualisierung komplexer Netzwerke und deren integrierter Daten. Verschiedenste Plugins erweitern die Funktionalität von *Cytoscape* für unterschiedliche Analysen, wie quantitative Petri-Netz, kinetische und topologische Analysen. Ein und dasselbe Anwendungsfenster wird nach und nach durch die Integration von Plugins erweitert. Das *Fern* Plugin ermöglicht die quantitative Analyse von SPNs, die u. a. in *SBML* definiert sind. Die Analyseergebnisse werden sowohl in Diagrammen zusammengefasst, als auch durch die Visualisierung der schaltenden Reaktion im Modell dargestellt. *Fern* wird auch zur kinetischen Analyse verwendet, um Modelle deterministisch und stochastisch zu analysieren. Die Plugins *CentiLib*, *CentiScaPe* und *BiNoM* werden zur topologischen Analyse von Zentralitäten und kürzesten Wegen verwendet. Die Modellstatistik wird bereitgestellt. Diese Maße können durch die *Brushing & Linking* Interaktionstechnik im Modell exploriert werden. *Cytoscape's* *CySBML* und *CySBGN* Plugins erlauben außerdem die Visualisierung von *SBML* Modellen in *SBGN*.

Außer der stöchiometrischen Analyse ermöglicht *Cytoscape* mit den vielen zur Verfügung stehenden Plugins (aktuell mehr als 200) vielfältige Möglichkeiten der Modell Analyse und Visualisierung. Obwohl *Cytoscape* als integriertes Programm erscheint, ist der Versuch, alle benannten Plugins zur metabolischen Modellierung integriert zu verwenden, gescheitert. Jedes der hier zu verwendenden Plugins wurde für eine spezielle *Cytoscape* Version entwickelt, ist für diese verfügbar und verifiziert. Das Plugin *Fern* ist nur für die Version 2.4 und *CentiLib* ist für die Version 2.6 geeignet. Von Version 2.7 bis Version 2.8 war *BiNoM* verwendbar und *CentiScaPe* findet ab Version 2.8 Anwendung. Die zum Import/Export und der Visualisierung notwendigen Plugins, wie *CySBML* und *CySBGN* stehen ab Version 2.8 zur Verfügung. Aufgrund dieser Versionsunterschiede ist bei der Evaluierung von *Cytoscape* keine funktionsfähige, integrierte Anwendung entstanden. Konkret konnte z.B. *Fern* nicht angewendet werden, wenn *CySBML* und *CySBGN* installiert waren. Die einzige Interaktion der getesteten Plugins konnte zwischen *CySBML* und *CySBGN* beobachtet werden. Eine Anpassung der erstellten Plugins für die neue *Cytoscape* Version 3 würde dieses Manko lösen und eine Basis für die gewünschte integrierte Anwendung zur metabolischen Modellierung darstellen.

Ähnlich zu *Cytoscape* ermöglicht das Visualisierungs- und Analysewerkzeug *VANTED* (*Visualization and Analysis of Networks containing Experimental Data*) [128] unter anderem die Erstellung von Netzwerken, die Integration verschiedener biologischer Datentypen und

die Exploration dieser Daten im Kontext von Netzwerken. Zudem finden verschiedene systembiologische Standards in der Visualisierung (*SBGN*) und dem Datenaustausch Anwendung (z. B. *SBML*, *PNML*, *SBGNML*). Als offenes System bietet VANTED eine Schnittstelle zur Erweiterung der Funktionalität durch Addons. Das CENTILIB Addon [64] wird zur topologischen, das FBASIMVIS Addon [62] zur stöchiometrischen und das PETRINET Addon [67] zur Petri-Netz Analyse verwendet. Die Visualisierung von *SBML* Modellen in *SBGN* wird durch das *SBGN-ED* Addon [32] erzielt.

VANTED ermöglicht die Analyse (außer die kinetische Analyse) von Modellen, die in spezifischen Formaten definiert sind. Im Gegensatz zu *SBW* und *Cytoscape* kommt VANTED mit einem Anwendungsfenster und ohne Versionsunterschiede der Addons aus. Als integrierte Anwendung zur metabolischen Modellierung fehlt VANTED die Implementierung der kinetischen Analyse und die Integration der vier Modellierungformalismen. Damit kann ein fließender und interaktiver Ablauf zur metabolischen Modellierung mit einer einheitlichen Visualisierung erzielt werden.

## 4.4 Erweiterte Anforderungen

Generell gibt es unter den Anwendungen und zugrunde liegenden Methodiken große Unterschiede in der Aufbereitung der Ergebnisse. Eine einheitliche Visualisierung für alle Modellierungformalismen ist empfehlenswert. Jedes der aufgelisteten Computerprogramme hat eigene Stärken und Schwächen und ist entweder spezialisiert auf einen dieser Formalismen oder umfasst mehrere. Durch Kombination verschiedener Computerprogramme können die Modellierungformalismen zur metabolischen Modellierung angewendet werden. Weiterhin können zusätzlich Visualisierungen eingesetzt werden. Obwohl die Anwendungen *SBW*, *Cytoscape* und VANTED bereits drei der vier geforderten Modellierungformalismen unterstützen, wird die Integration und Interaktion der möglichen Erweiterungen vermisst. Die Anforderungen an eine integrierte Anwendung, die (1) kinetische, Petri-Netz, stöchiometrische und topologische Modellierung, (2) einheitliche Visualisierung (*SBGN*), (3) Analyseergebnisse im Kontext des Modells und (4) Exploration durch Interaktionstechniken unterstützt, konnte in dieser Evaluierung nicht ermittelt werden. Nach Dandekar und Koautoren [33] werden die Anforderungen, Visualisierung und die Verwendung verschiedener Modellierungformalismen als Auswahlkriterien einer geeigneten Anwendung zur metabolischen Modellierung gesetzt. Die folgenden Kapitel beschreiben die Entwicklung einer Methodik, die diese Anforderungen erfüllt (in Kapitel 5) und somit eine Realisierung in Form einer integrierten Anwendung ermöglicht (in Kapitel 6).

## 5 Methodik

Die Systembiologie zieht ihren Vorteil aus Computerwerkzeugen und experimentellen Daten aus Hochdurchsatztechnologien, um biologische Prozesse zu modellieren [106]. Verschiedene Modellierungsformalismen werden bereits für metabolische Modelle verwendet, die auf unterschiedliche Art und Weise den Metabolismus eines Organismus oder eines Ausschnittes dessen abstrahieren. Unter anderem werden topologische, stöchiometrische, Petri-Netz und kinetische Modelle rekonstruiert (siehe Kapitel 3). Die Wahl des Modellierungsformalismus ist von den vorhandenen detaillierten Experimentaldaten und dem Zweck, welchen das Modell erfüllen soll, abhängig. Die verschiedenen Formalismen werden von unterschiedlichen Computerwerkzeugen unterstützt (siehe Kapitel 4 Tabelle 4.5).

Zur Integration unterschiedlicher Modellierungsformalismen metabolischer Modelle wird in diesem Kapitel eine Methodik erarbeitet. Vier etablierte Formalismen werden zur integrativen Analyse metabolischer Modelle angewendet und bringen strukturelle und dynamische Eigenschaften hervor, die ergänzende Erkenntnisse liefern. Zusätzlich führt die visuelle Integration ermittelter Analyseergebnisse in entsprechende Elemente des metabolischen Modells zur Kombination von Modellstruktur und -dynamik. Geeignete Interaktionstechniken unterstützen die Exploration und Interpretation erhobener Ergebnisse und führen zu einem umfassenden Verständnis des zugrunde liegenden biologischen Systems. Das Vorgehen in diesem Kapitel ist in Abbildung 5.1 dargestellt und wird im Folgenden erläutert.

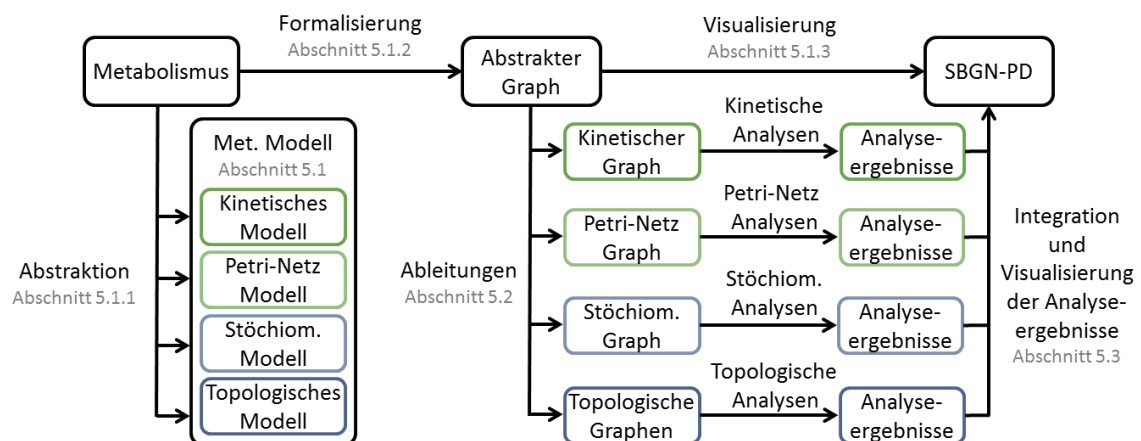


Abbildung 5.1: Struktur des Kapitels Methodik.

### 5.1 Metabolisches Modell

Zunächst wird die Abstraktion vom Metabolismus anhand eines abstrakten metabolischen Modells beschreiben und im Rahmen verschiedener Modellierungsformalismen erläutert. Danach erfolgt die Formalisierung des metabolischen Modells unter Verwendung einer

einheitlichen und erweiterbaren Datenstruktur. Abschließend wird die standardisierte Visualisierung des metabolischen Modells vorgestellt.

### 5.1.1 Abstraktion

Die in der vorliegenden Arbeit verwendeten metabolischen Modelle beschreiben verschiedene topologische, stöchiometrische, Petri-Netz und kinetische mathematische Modelle (siehe Kapitel 3 zur Übersicht). Diese unterschiedlichen mathematischen Beschreibungen werden in ein abstraktes metabolisches Modell wie folgt übertragen.

#### 5.1.1.1 Strukturelle Beschreibung

Jedes dieser Modelle (außer den speziellen topologischen Modellen, *Metabolit-* und *Reaktionsgraph*, siehe Kapitel 3 Abschnitt 3.5) besteht strukturell aus einer Menge von Metaboliten und einer Menge biochemischer Reaktionen. Zwischen diesen beiden Mengen existieren Beziehungen. Eine biochemische Reaktion muss darin mit mindestens einem Metaboliten verbunden sein. Ein minimales metabolisches Modell besteht somit aus einer Reaktion, verbunden mit einem Metaboliten (Substrat oder Produkt einer Reaktion).

Jeder Metabolit besitzt eine Lokalisation (z. B. Kompartiment einer Zelle) und kann durch verschiedene Beziehungen in eine biochemische Reaktion als Substrat, Produkt und/oder Inhibitor eingehen. Es ist möglich, dass ein Substrat oder Produkt gleichzeitig auch Inhibitor einer biochemischen Reaktion ist. Es handelt sich dann um eine Substrat- oder Produktinhibition. Eine biochemische Reaktion kann gleichzeitig durch ein Substrat und ein Produkt inhibiert werden. Stöchiometrische, Petri-Netz und kinetische Modelle beinhalten zusätzlich Stöchiometrien, die quantitative Beziehungen zwischen Substraten und Produkten einer Reaktion angeben.

Des Weiteren kann eine biochemische Reaktion entweder irreversibel oder reversibel sein, wobei diese Eigenschaft durch die jeweiligen Beziehungen zwischen Metaboliten und Reaktionen festgelegt wird. Eine reversible Reaktion wird abhängig vom Reaktionsmechanismus, bereits von Substraten in der Rückreaktion und von Produkten in der Hinreaktion, inhibiert. In der kinetischen Modellierung finden diese Inhibitionen (impliziert durch die kinetische Ratengleichung) Verwendung.

Biochemische Reaktionen können spontan- oder enzymkatalysiert ablaufen oder den Transport von Metaboliten über die Kompartiment-, als auch die Systemgrenze (Import, Export) übernehmen (Transportreaktionen). Die meisten Reaktionen eines metabolischen Modells sind enzymkatalysiert, einige sind Transportreaktionen. Spontane Reaktionen finden sehr selten statt.

Die verwendete Struktur eines abstrakten metabolischen Modells ist hiermit beschrieben. Die Struktur eines speziellen topologischen Modells ist hingegen verschieden von der hier erläuterten Struktur. Aber diese *Metabolit-* und *Reaktionsgraphen* können aus dem abstrakten metabolischen Modell abgeleitet werden (siehe Kapitel 3 Abschnitt 3.5). Diese Graphen bestehen aus entweder Metaboliten oder Reaktionen und Beziehungen zwischen diesen. Somit bilden diese Graphen nicht die Basis der Strukturbeschreibung eines abstrakten metabolischen Modells.

#### 5.1.1.2 Funktionelle Beschreibung

Verschiedene Modellierungsformalismen erfordern zur Analyse zusätzlich die Integration von detaillierten Informationen. Deshalb werden für biochemische Reaktionen kinetische

Ratengleichungen oder gewichtete stochastische Schaltraten definiert, die die quantitative kinetische und Petri-Netz (eSPN) Analyse erlauben. Des Weiteren erhalten Metabolite eine initiale Mengenangabe (Konzentration oder Molekülanzahl), wobei zur Petri-Netz Analyse zusätzlich eine Kapazität der Metabolite angegeben werden kann, sodass eine bestimmte Menge eines Metaboliten nicht überschritten werden darf. Zur stöchiometrischen Analyse werden Einschränkungen der Reaktionskapazität (Flusskapazität) erforderlich. Dazu werden untere- und obere Schranken der Reaktionsraten (minimaler und maximaler metabolischer Fluss einer Reaktion) festgelegt. Zudem wird eine Reaktion als Zielfunktion festgelegt, die in der Optimierung verwendet wird.

Das abstrakte metabolische Modell ist mit allen notwendigen Informationen, um für die vier Modellierungsformalismen verwendbar zu sein, definiert. Im Folgenden wird nun die Formalisierung dieses Modells erarbeitet.

### 5.1.2 Formalisierung

Die Formalisierung des abstrakten metabolischen Modells erfolgt in Form einer geeigneten Datenstruktur, die die Beziehungen zwischen der Metabolit- und Reaktionsmenge darstellt. Der Graphformalismus (siehe Kapitel 2 Abschnitt 5.2) ist für diesen Zweck eine geeignete Datenstruktur.

**Vorüberlegung** Es ergeben sich im Folgenden zwei mögliche Vorgehensweisen:

1. Erstellung von Datenstrukturen (spezielle Graphen) für den entsprechenden Modellierungsformalismus, wobei jeder Graph einheitlich visualisiert wird.
2. Entwicklung einer einheitlichen Datenstruktur (*abstrakter Graph*), die das Ableiten der speziellen Graphen für die verschiedenen Modellierungsformalismen erlaubt und einheitlich visualisiert wird.

**Zu 1.:** Die erste Vorgehensweise bedarf der Definition von mindestens vier speziellen Graphen. Zur kinetischen Modellierung eignet sich ein gerichteter, attributierter, bipartiter Graph (Definition siehe Kapitel 2 Abschnitt 5.2), der durch reversible Kanten, irreversible Kanten und Inhibitoranten die Elemente der Metabolit- und Reaktionsmenge verbindet. Zusätzlich werden Abbildungen benötigt, die den Reaktionsknoten kinetische Reaktionsraten und den Metabolitknoten Konzentrationen zuordnen.

Die Petri-Netz Modellierung hingegen erfordert einen gerichteten, attributierten, bipartiten Graphen, der nur irreversible Kanten und Inhibitoranten beinhaltet. Hier werden Abbildungen benötigt, die den Reaktionsknoten stochastische Schaltraten und den Metabolitknoten Konzentrationen zuordnen.

In der stöchiometrischen Modellierung werden nur reversible und irreversible Kanten zur Verbindung der beiden Knotenmengen Metabolite und Reaktionen verwendet. Zusätzlich werden für Reaktionsknoten Abbildungen notwendig, die untere- und obere Schranken und eine Zielfunktion zur Optimierung festlegen. Auch hier wird ein gerichteter, attributierter, bipartiter Graph definiert.

Die topologische Modellierung verwendet hauptsächlich gerichtete, attributierte Graphen, den *Metabolit-* oder *Reaktionsgraphen*. Diese unipartiten Graphen (siehe Kapitel 2 Abschnitt 5.2) besitzen entweder Metabolite oder Reaktionen, die miteinander verbunden sind.

Diese fünf verschiedenen Graphen sind nicht isomorph zueinander (siehe Kapitel 2 Abschnitt 5.2). Es werden bipartite Graphen zur kinetischen, Petri-Netz und stöchiometrischen Modellierung verwendet und unipartite Graphen zur topologischen Modellierung. Zudem gibt es Unterschiede zwischen den bipartiten Graphen. Der *Petri-Netz Graph* verwendet nur irreversible Kanten und Inhibitorkanten. Der *kinetische Graph* verwendet zusätzlich reversible Kanten und der *stöchiometrische Graph* nutzt nur reversible und irreversible Kanten.

Eine gegenseitige Umwandlung dieser Graphen ist ohne großen Aufwand nicht möglich. Es bedarf vieler verschiedener Ableitungen (vom *kinetischen Graphen* zum *Petri-Netz Graphen* und umgekehrt). Eine einheitliche Visualisierung ist unter der Betrachtung der *topologischen Graphen* (unipartite Graphen) nicht möglich.

**Zu 2.:** Die zweite Vorgehensweise ermöglicht die Verwendung einer einheitlichen Datenstruktur, die *abstrakter Graph* genannt wird. Der *abstrakte Graph* ist ein gerichteter, attributierter, bipartiter Graph, der eine einheitliche Visualisierung zulässt. Je nach Bedarf wird eine Umformung in die speziellen Graphen (*kinetischer, Petri-Netz, stöchiometrischer* und *topologische Graphen*) vorgenommen, um die speziellen Analysen zu verwenden. Die Visualisierung des metabolischen Modells bleibt dabei erhalten. Verschiedene Analyseergebnisse werden in die einheitliche Visualisierung integriert und im Kontext des metabolischen Modells, mittels Interaktionstechniken, exploriert. Diese Vorgehensweise erfordert fünf Ableitungen vom *abstrakten Graphen* zum 1) *kinetischen Graphen*, 2) *Petri-Netz Graphen*, 3) *stöchiometrischen Graphen*, 4) *topologischen Metabolitgraphen* und 5) *topologischen Reaktionsgraphen*. Eine weitere Ableitung vom *abstrakten Graphen* zur einheitlichen Visualisierung wird durchgeführt. Weiterhin erlauben verschiedene Datenzuordnungsfunktionen eine einheitliche Visualisierung der Analyseergebnisse im metabolischen Modell.

Die zweite Vorgehensweise erfordert weniger Aufwand und ist ein genereller Ansatz. Der *abstrakte Graph* kann um weitere Modellierungsformalismen erweitert werden. Zusätzliche Analysen werden durch die Erweiterung der Attributmenge und das Hinzufügen einer zusätzlichen Ableitungsfunktion des *abstrakten Graphen* ermöglicht.

**Abstrakter Graph** Zunächst wird der *abstrakte Graph*, entsprechend der Anforderungen in Abschnitt 5.1.1, definiert. Ein *abstrakter Graph* eignet sich als einheitliche Datenstruktur, welche die Speicherung und Ableitung von speziellen metabolischen Modellen (*speziellen Graphen*) ermöglicht. Der *abstrakte Graph*  $G_{\text{Abstrakt}} = (M, R, K, A)$  ist ein gerichteter, attributierter, bipartiter Graph. Dieser Graph besteht aus den beiden nicht-leeren, endlichen Knotenmengen Metabolite  $M$  und biochemischen Reaktionen  $R$  eines metabolischen Modells, wobei beide Mengen disjunkt sind,  $M \cap R = \emptyset$ . Zudem enthält der *abstrakte Graph* eine endliche Menge gerichteter Kanten  $K \subseteq (M \times R) \cup (R \times M)$  und eine endliche Menge von Attributen  $A = \{\text{Typ}, \text{Stöchiometrie}, \text{Lokalisation}, \text{Bezeichner}, \text{Menge}, \text{Kapazität}, \text{Rate}, \text{Schranken}, \text{Zielfunktion}\}$ , die mit den im Folgenden definierten Abbildungen Knoten und Kanten zugeordnet werden:

- $\text{Typ} : K \rightarrow \{vi, pi, vr, pr, i\}$  ordnet jeder Kante einen Typ ( $vi$  Verbrauch irreversibel,  $pi$  Produktion irreversibel,  $vr$  Verbrauch reversibel,  $pr$  Produktion reversibel oder  $i$  Inhibition) zu. Eine gerichtete Kante von einem Metaboliten zu einer Reaktion kann vom Typ  $vi$ ,  $vr$  oder  $i$  sein (d. h.  $\forall k \in (M \times R) : \text{Typ}(k) = vi \vee \text{Typ}(k) = vr \vee \text{Typ}(k) = i$ ). Verläuft eine gerichtete Kante von einer Reaktion zu einem Metaboliten, kann diese



entweder vom Typ  $pi$  oder  $pr$  sein (d. h.  $\forall k \in (R \times M) : Typ(k) = pi \vee Typ(k) = pr$ ). Reversible Kanten werden zur besseren Unterscheidung zu den irreversiblen Kanten durch Doppelpfeile dargestellt, wobei die Hauptrichtung vom Substrat (verbrauchter Metabolit) zum Produkt (produzierter Metabolit) durch eine schwarze Pfeilspitze gekennzeichnet ist.

- *Stöchiometrie* :  $K' \rightarrow \mathbb{R}_{>0}$  ordnet jeder Kante vom Typ  $vi$ ,  $vr$ ,  $pi$  oder  $pr$  aus der Menge  $K' = \{k \in K \mid \neg(Typ(k) = i)\}$  eine reelle Zahl größer als 0 (Stöchiometrie) zu.
- *Bezeichner* :  $M \cup R \rightarrow \Sigma^*$  ordnet jedem Metaboliten und jeder Reaktion ein Wort über dem Alphabet (Bezeichner) zu.
- *Lokalisation* :  $M \rightarrow \Sigma^*$  ordnet jedem Metaboliten ein Wort über dem Alphabet (Lokalisation) zu.
- *Kapazität* :  $M \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$  ordnet jedem Metaboliten eine nicht-negative reelle Zahl (Kapazität), die auch unbeschränkt sein kann  $\{\infty\}$ , zu.
- *Menge* :  $M \rightarrow \mathbb{R}_{\geq 0}$  ordnet jedem Metaboliten eine nicht-negative reelle Zahl (Konzentration des Metaboliten) zu. Zudem muss diese Menge der Metabolite kleiner oder gleich der Kapazität des Metaboliten sein,  $\forall m \in M : Menge(m) \leq Kapazität(m)$ .
- *Rate* :  $R \rightarrow \{\{h, j\}, h, j, \{\}\}$  ordnet jeder Reaktion eine kinetische Ratengleichung  $j \in J$  aus der Menge aller Reaktionskinetiken oder eine nicht-negative reelle Zahl (stochastische Schaltrate)  $h \in \mathbb{R}_{\geq 0}$  oder die leere Menge  $\{\}$  zu.
- *Schranken* :  $R \rightarrow (min, max)$  mit  $min, max \in \mathbb{R}_{\geq 0}$  und  $min \leq max$  ordnet jeder Reaktion ein geordnetes Paar nicht-negativer reeller Zahlen (untere und obere Schranke) zu. Die untere Schranke muss kleiner oder gleich der oberen Schranke sein.
- *Zielfunktion* :  $R \rightarrow \{0, 1\}$ , mit  $\forall r, r' \in R : Zielfunktion(r) = 1 \wedge Zielfunktion(r') = 1 \Rightarrow r = r'$  ordnet jeder Reaktion einen Wert 0 oder 1 zu, wobei nur eine Reaktion den Wert 1 (Zielfunktion für die Optimierung) erhält.

Des Weiteren seien die folgenden Bedingungen erfüllt:

Für alle Reaktionen  $r \in R$  gilt: 1) es existiert mindestens eine eingehende und eine ausgehende Kante (wobei die eingehende Kante nicht vom Typ  $i$  ist) und 2) wenn eine eingehende oder ausgehende Kante vom Typ reversibel (irreversibel) ist, dann sind alle eingehenden und ausgehenden Kanten reversibel (irreversibel). Mit dieser Regel ist eine Reaktion entweder mit reversiblen oder irreversible Kanten verbunden und keine Mischung dieser.

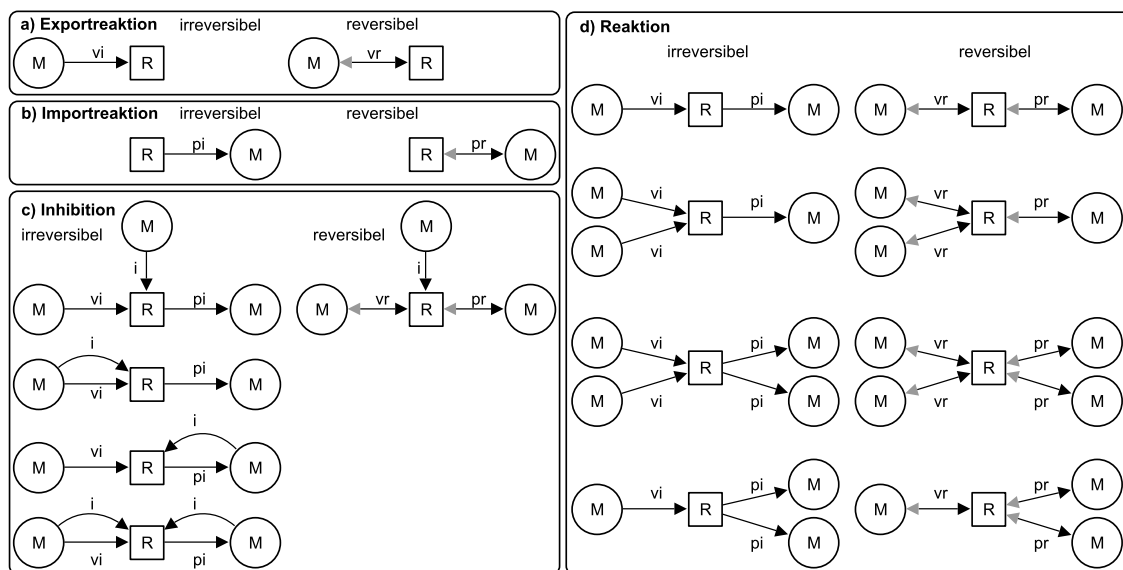
Zwischen einem Metaboliten  $m \in M$  und einer Reaktion  $r \in R$  können maximal zwei gerichtete Kanten  $k, k' \in K$  unterschiedlichen Typs existieren. Wenn zwei Kanten  $k$  und  $k' m$  mit  $r$  verbinden, dann ist  $Typ(k) = vi$  und der  $Typ(k') = i$ . In diesem Fall kann ein Metabolit Substrat als auch Inhibitor einer Reaktion sein, um eine Substratinhibition der Reaktion bei zu hoher Substratkonzentration (siehe Abschnitt 2.1) darzustellen.

Verbindet eine Kante  $k r$  mit  $m$  und eine weitere Kante  $k' m$  mit  $r$ , dann ist  $k$  vom Typ  $pi$  und der Typ von  $k'$  ist  $i$ . Hier handelt es sich um eine Produktinhibition. Der Metabolit ist Produkt und gleichzeitig Inhibitor einer Reaktion. Bei reversiblen Reaktionen sind diese beiden Inhibitionen bereits durch den Reaktionsmechanismus vorgegeben und werden an dieser Stelle nicht explizit formuliert.

Außerdem seien die folgenden Mengen definiert. Die Kantenmenge  $K$  wird abhängig vom Typ Attribut aus drei Teilmengen gebildet:  $K = K_i \cup K_{ir} \cup K_r$ . Die Teilmenge der Inhibitorikanten ist  $K_i = \{k \in K | Typ(k) = i\}$ , die Teilmenge der irreversiblen Kanten ist  $K_{ir} = \{k \in K | Typ(k) = vi \vee Typ(k) = pi\}$  und die Teilmenge der reversiblen Kanten ist  $K_r = \{k \in K | Typ(k) = vr \vee Typ(k) = pr\}$ . Die Menge der Metabolite  $M$  besteht aus einer Teilmenge von Metaboliten  $M_{vp}$ , die in Reaktionen entweder verbraucht oder produziert werden  $M_{vp} = \{m \in M | \exists r \in R : (m, r) \in K_r \vee (m, r) \in K_{ir}\} \cup \{m' \in M | \exists r \in R : (r, m') \in K_r \vee (r, m') \in K_{ir}\}$ .

Des Weiteren besitzt der *abstrakte Graph* Datenzuordnungsfunktionen  $zn$  und  $zg$  zur Integration struktureller und dynamischer Analyseergebnisse in Knoten und Kanten (siehe Abschnitt 5.3.1).

Aufgrund der Definition des *abstrakten Graphen* entsteht eine Datenstruktur, die in Form eines Graphen aus den in Abbildung 5.2 dargestellten Grundelementen besteht.

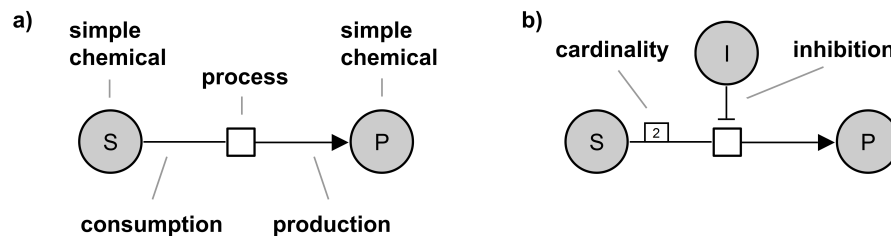


**Abbildung 5.2:** Datenstruktur des abstrakten metabolischen Modells: bipartiter, attributierter, gerichteter Graph. Dieser *abstrakte Graph* repräsentiert a) Export- und b) Importreaktionen, c) Inhibition von Reaktionen und d) verschiedene Reaktionen (links irreversibel, rechts reversibel). Die Bestandteile des Graphen sind Metabolitknoten (runde Symbole, Bezeichner  $M$ ) und Reaktionsknoten (eckige Symbole, Bezeichner  $R$ ) verbunden durch gerichtete Kanten (Linien), die verschiedene Symbole an den Endpunkten besitzen (Pfeil an einem oder beiden Endpunkten). Zudem erhalten die Kanten Attribute u. a. das Typ Attribut ( $vi/vr =$  ir-/reversibler Verbrauch,  $pi/pr =$  ir-/reversible Produktion,  $i =$  Inhibition). Reversible Kanten werden zur besseren Unterscheidung zu den irreversiblen Kanten durch Doppelpfeile dargestellt, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist.

### 5.1.3 Visualisierung

Die einheitliche Datenstruktur des metabolischen Modells, der *abstrakte Graph*, erlaubt spezielle Datenstrukturen (*spezielle Graphen*) abzuleiten. Diese *speziellen Graphen* ermöglichen verschiedene Modellierungsfomalismen anzuwenden. Die graphische Repräsentation metabolischer Modelle kann für jeden Modellierungsfomalismus verschieden sein. Eine einheitliche Visualisierung wird unabhängig von dem Grad enthaltener Details mittels *SBGN* [100] erreicht (ausführliche Erläuterung siehe Kapitel 2 Abschnitt 2.3.3.1).

Aufgrund der Definition eines *abstrakten Graphen* (siehe Abschnitt 5.1.2), wird eine einheitliche Visualisierung trotz Verwendung verschiedener Modellierungsformalismen möglich. Die zwei attributierten Knotenmengen Metabolite  $M$  und biochemische Reaktionen  $R$  sowie die durch Attribute definierte Menge gerichteter Kanten  $K$  können in *SBGN-PD* standardisiert visualisiert werden. In *SBGN-PD* gibt es u. a. eine Knotenmenge *entity pool nodes*, die verschiedene graphische Symbole (engl. glyphs) für unterschiedliche Objekte enthält, wie z. B. das Symbol *simple chemical* für Metabolite. Eine weitere Knotenmenge *process nodes* enthält verschiedene Symbole zur Darstellung unterschiedlicher Prozesse, wie z. B. das Symbol *process* für biochemische Reaktionen. Die Kantenmenge in *SBGN-PD* wird *connecting arcs* genannt. Kanten verbinden u. a. Elemente der Menge *entity pool nodes* mit Elementen der Menge *process nodes* und umgekehrt. Verschiedene Symbole an den Endpunkten einer Kante veranschaulichen die Beziehung zwischen diesen Elementen. Zum Beispiel kann ein *simple chemical* durch die Kante *consumption* mit einem *process* und dieser wiederum durch die Kante *production* mit einem weiteren *simple chemical* verbunden sein. Ein Metabolit wird somit in einer Reaktion in einen anderen Metaboliten umgewandelt, siehe Abbildung 5.3 a).



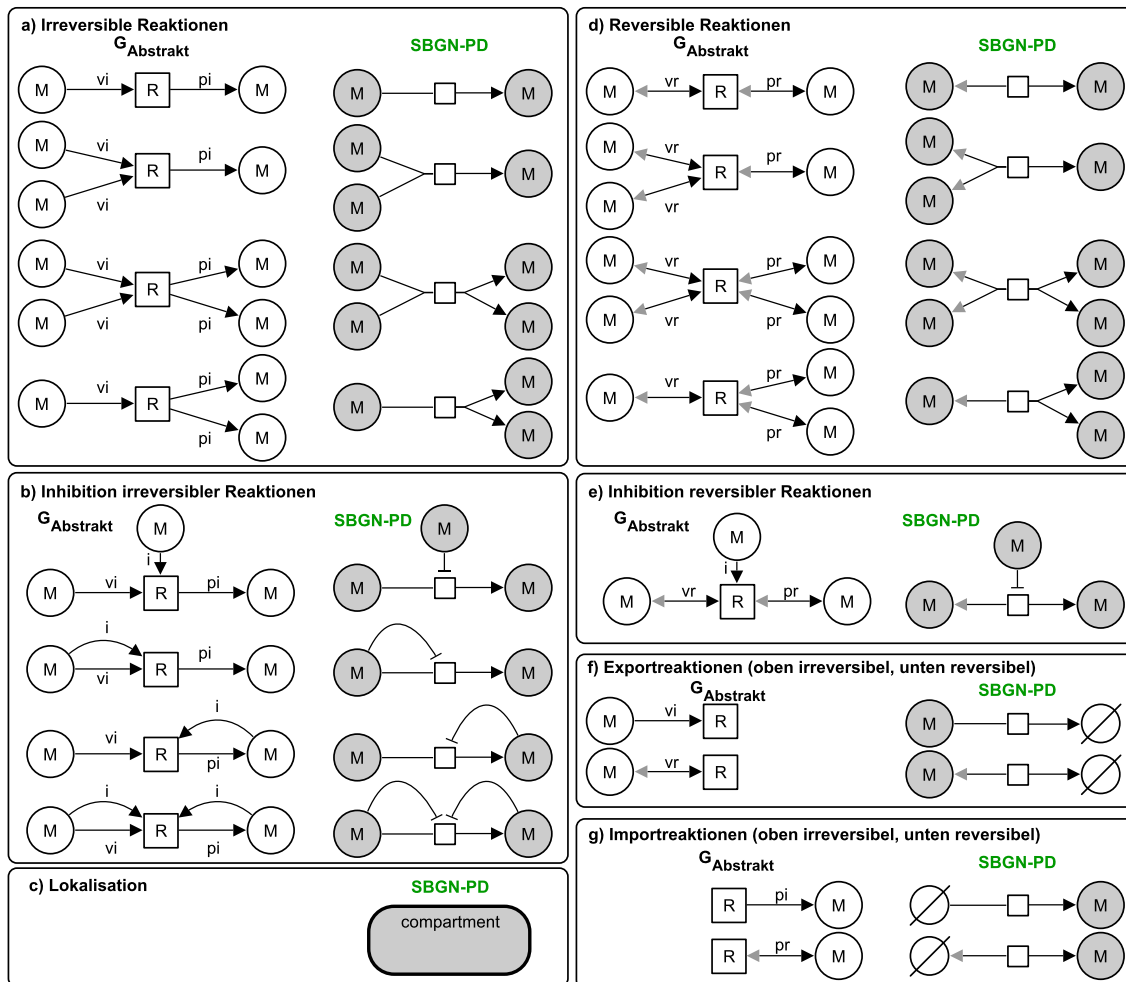
**Abbildung 5.3:** Visualisierung in *SBGN-PD*: a) irreversible biochemische Reaktion, b) inhibierte irreversible biochemische Reaktion mit der Angabe eines stöchiometrischen Koeffizienten (*cardinality*) des Metaboliten  $S$ .

Die Ableitung des *abstrakten Graphen*  $G_{Abstrakt} = (M, R, K, A)$  in eine *SBGN-PD* Visualisierung erfolgt nach folgendem Schema:

- Elemente der Menge Metabolite  $m \in M$  werden in Form von *simple chemicals*  $\in$  *entity pool nodes* visualisiert.
- Elemente der Menge Reaktionen  $r \in R$  werden in Form von *process*  $\in$  *process nodes* visualisiert.
- Die gerichteten Kanten  $k \in K$  zwischen Metaboliten und Reaktionen und umgekehrt werden aufgrund der Typ Attribute verschieden visualisiert:
  - Kanten vom Typ Verbrauch irreversibel  $vi$  werden zu *consumption*  $\in$  *connecting arcs*,
  - Kanten vom Typ Produktion irreversibel  $pi$  werden zu *production*  $\in$  *connecting arcs*,
  - Kanten vom Typ Verbrauch reversibel  $vr$  werden zu *production*  $\in$  *connecting arcs* in umgekehrter Richtung,
  - Kanten vom Typ Produktion reversibel  $pr$  werden zu *production*  $\in$  *connecting arcs*,
  - Kanten vom Typ Inhibition  $i$  werden zu *inhibition*  $\in$  *connecting arcs*, siehe Abbildung 5.3 b).

- Das Kanten Attribut *Stöchiometrie* wird in Form der *cardinality* einer Kante visualisiert, siehe Abbildung 5.3 b).
- Das Metabolit Attribut *Lokalisation* wird zu einem neuen Symbol dem *compartment*, der als Container die Menge der Metabolite enthält. Die Lokalisation der Reaktionen ist vom *compartment* unabhängig. Eine Reaktion kann sich innerhalb, außerhalb oder auf der Begrenzung des *compartments* befinden.
- Die Import- oder Exportreaktionen werden in SBGN-PD mit einem zusätzlichen Symbol *source and sink*  $\in$  *entity pool nodes* visualisiert, sodass beiderseits eines *process* Knotens Metabolite vorhanden sind.
- Das Knoten Attribut *Bezeichner* wird in Form des *labels* den Knoten hinzugefügt.

Insgesamt entsteht eine einheitliche Visualisierung für den *abstrakten Graphen*, der das metabolische Modell repräsentiert, siehe Abbildung 5.4.



**Abbildung 5.4:** Visualisierung des *abstrakten Graphen* in SBGN-PD (links *abstrakter Graph*, rechts entsprechende SBGN-PD Visualisierung): a) Irreversible Reaktionen, b) Inhibition irreversibler Reaktionen, c) Lokalisation (*compartment*) der Metabolitknoten (Zugehörigkeit durch Färbung gekennzeichnet), d) Reversible Reaktionen, e) Inhibition reversibler Reaktionen, f) Exportreaktionen (oben irreversibel, unten reversibel) und g) Importreaktionen (oben irreversibel, unten reversibel).

Die an dieser Stelle verwendete *SBGN-PD* Symbolmenge beschreibt einen kleinen Teil des Möglichen. Im Prinzip können zur Visualisierung von weiteren Details oder für die Repräsentation von unspezifizierten Molekülen und Reaktionen weitere Symbole verwendet werden (siehe Kapitel 2 Abschnitt 2.3.3.1).

## 5.2 Ableitung spezieller Modelle zur Analyse

### 5.2.1 Kinetisches Modell

Ein dynamisches metabolisches Modell wird als kinetisches Modell bezeichnet, wenn Reaktionsmechanismen durch detaillierte kinetische Informationen (Ratengleichungen, Parameter und Metabolitkonzentrationen) beschrieben sind.

#### 5.2.1.1 Graph

Das kinetische Modell wird in Form eines *kinetischen Graphen*  $G_{Kinetik}$ , aus dem *abstrakten Graphen*  $G_{Abstrakt}$ , abgeleitet. Der *kinetische Graph*  $G_{Kinetik} = (M_K, R_K, K_K, A_K)$  ist ein gerichteter, attributierter, bipartiter Graph. Die beiden endlichen Knotenmengen bestehen aus Metaboliten  $M_K$  sowie Reaktionen  $R_K$  und sind disjunkte Mengen,  $M_K \cap R_K = \emptyset$ . Weiterhin enthält dieser Graph eine endliche Kantenmenge gerichteter Kanten  $K_K \subseteq (M_K \times R_K) \cup (R_K \times M_K)$  und eine endliche Menge von Attributen  $A_K = \{Typ, Gewicht, Lokalisation, Bezeichner, Menge, Rate\}$ . Der  $G_{Kinetik}$  entspricht außer einigen Attributen bereits dem  $G_{Abstrakt}$ , sodass die Ableitung des  $G_{Kinetik}$  aus  $G_{Abstrakt}$  wie folgt durchgeführt wird:

- $M_K = S$  die Menge der Metabolite in  $G_{Kinetik}$  entspricht der Menge der Metabolite in  $G_{Abstrakt}$ ,
- $R_K = R$  die Menge der Reaktionen in  $G_{Kinetik}$  entspricht der Menge der Reaktionen in  $G_{Abstrakt}$ ,
- $K_K = K$  die Menge der Kanten in  $G_{Kinetik}$  entspricht der Menge der Kanten in  $G_{Abstrakt}$ ,
- $A_K \subseteq A$  die Menge der Attribute in  $G_{Kinetik}$  entspricht einer Teilmenge der Attribute in  $G_{Abstrakt}$ ,  $A_K = \{Typ, Stöchiometrie, Lokalisation, Bezeichner, Menge, Rate\}$ , wobei die *Rate* nur aus der Menge der Reaktionskinetiken  $J$  oder der leeren Menge  $\{\}$  (im Falle keine Reaktionskinetik oder nur stochastische Schaltrate wurden vergeben) besteht.

Diese Ableitung resultiert in keinen strukturellen Unterschieden zwischen *kinetischem Graph* und *abstrakten Graphen*, aber die abgeleitete Attributmeng ist geringer.

#### 5.2.1.2 Analysen

In der kinetischen Analyse metabolischer Modelle (Details siehe Kapitel 3 Abschnitt 3.2) wird der *kinetische Graph* in gewöhnliche Differentialgleichungen (*ODEs*) übersetzt. Diese *ODEs* werden für jeden Zeitschritt mit den aktualisierten Metabolitkonzentrationen gelöst. Das Ergebnis der kinetischen Analyse sind Metabolitkonzentrationen und Reaktionsraten über die Zeit.

## 5.2.2 Petri-Netz Modell

Metabolische Modelle können auch in Form eines Petri-Netzes repräsentiert werden. Petri-Netze ermöglichen metabolische Modelle strukturell als auch dynamisch zu betrachten.

### 5.2.2.1 Graph

Der bereits in Kapitel 3 Abschnitt 3.3 eingeführte erweiterte Petri-Netz Formalismus (für  $ePNs$  und  $eSPNs$ ) kann in Form eines *Petri-Netz Graphen*  $G_{Petri-Netz}$  aus  $G_{Abstrakt}$  abgeleitet werden. Hier liegt der Schwerpunkt in der Aufspaltung reversibler Reaktionen in eine irreversible Hin- und eine irreversible Rückreaktion, da Petri-Netze keine reversiblen Reaktionen beinhalten.

Der Graph  $G_{Petri-Netz} = (M_P, R_P, K_P, A_P)$  ist ein gerichteter, attributierter, bipartiter Graph, der aus den beiden endlichen Knotenmengen Metabolite  $M_P$  und Reaktionen  $R_P$  eines metabolischen Modells besteht, wobei beide Mengen disjunkt sind,  $M_P \cap R_P = \emptyset$ . Zudem enthält ein *Petri-Netz Graph* eine endliche gerichtete Kantenmenge  $K_P \subseteq (M_P \times R_P) \cup (R_P \times M_P)$  und eine endliche Menge von Attributen  $A = \{Typ, Stöchiometrie, Lokalisation, Bezeichner, Menge, Kapazität, Rate\}$ . Diese Mengen werden aus dem *abstrakten Graphen* wie folgt abgeleitet:

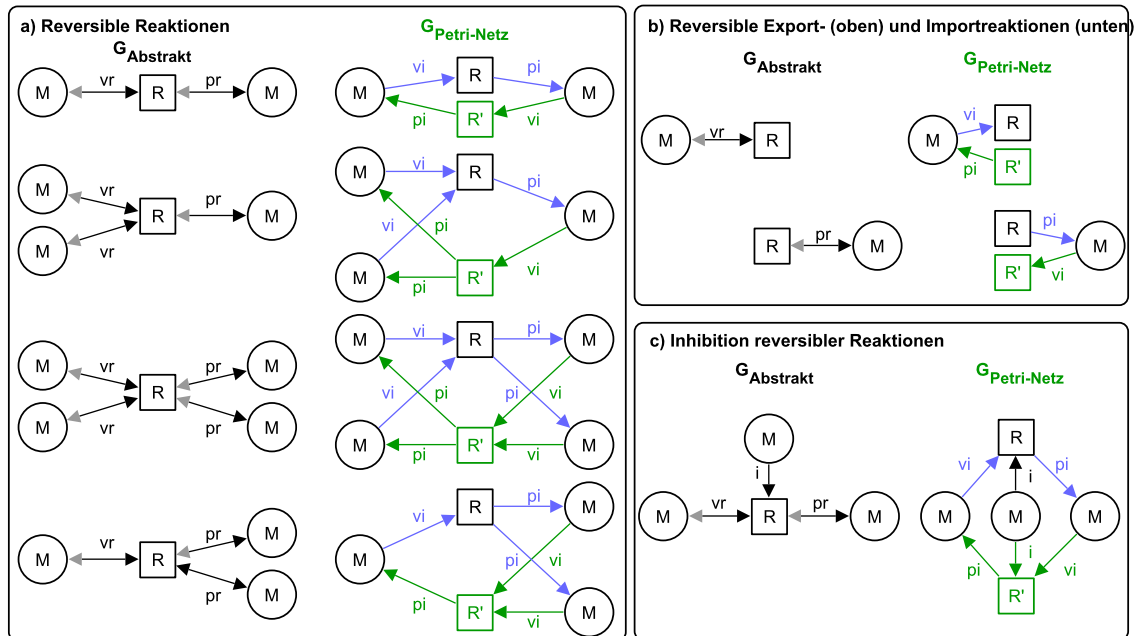
- $M_P = M$  die Menge der Metabolite in  $G_{Petri-Netz}$  entspricht der Menge der Metabolite in  $G_{Abstrakt}$ .
- $R_P = R \cup R_{P,neu}$ , die Menge der Reaktionen in  $G_{Petri-Netz}$  entspricht der Menge der Reaktionen in  $G_{Abstrakt}$  und zusätzlichen Reaktionsknoten. Denn reversible Reaktionen werden in je eine irreversible Hin- und eine irreversible Rückreaktion umgewandelt, siehe Abbildung 5.5. Dabei wird jeder Reaktionsknoten einer reversiblen Reaktion ( $r$ , auch Ursprungsreaktion genannt) zur Hinreaktion und ein neuer Reaktionsknoten zur Rückreaktion ( $r_{neu}$ ) hinzugefügt. Die Menge der Rückreaktionsknoten  $R_{P,neu}$  ist wie folgt definiert:  $R_{P,neu} = \{r_{P,neu} | \exists r \in R : ((m, r) \in K_r \vee (r, m') \in K_r) \wedge m, m' \in M\}$ . Zusätzlich wird den neuen Rückreaktionen die Rate der Ursprungsreaktionen und deren Bezeichner mit zusätzlichem Vermerk ( $'$ ) zugeordnet.
- $K_P = K_{ir} \cup K_i \cup K_{P,neu}$  die Menge der Kanten in  $G_{Petri-Netz}$  entspricht den Kantenteilmengen  $K_{ir}$  und  $K_i$  des  $G_{Abstrakt}$  und einer Menge zusätzlicher Kanten  $K_{P,neu}$ . Aufgrund der Umwandlung reversibler Reaktionen werden für die Hin- und Rückreaktionen irreversible Verbrauchs- ( $K_{P,vi}$ ) und Produktionskanten ( $K_{P,pi}$ ) als auch Inhibitionskanten ( $K_{P,i}$ ) der Kantenmenge hinzugefügt:  $K_{P,neu} = K_{P,vi} \cup K_{P,pi} \cup K_{P,i}$ , siehe Abbildung 5.5.

Abhängig von den reversiblen Ursprungsreaktionen  $r$  werden die Kantenteilmengen wie folgt gebildet: Die Teilmenge der Verbrauchskanten der Hin- und Rückreaktion ist  $K_{P,vi} = \{(m, r) \in K_r | m \in M \wedge r \in R\} \cup \{(m', r_{P,neu}) | \exists r \in R : (r, m') \in K_r \wedge m' \in M\}$ , die der Produktionskanten ist  $K_{P,pi} = \{(r, m') \in K_r | m' \in M \wedge r \in R\} \cup \{(r_{P,neu}, m) | \exists r \in R : (m, r) \in K_r \wedge m \in M\}$  und die Inhibitionskanten werden mit der Teilmenge  $K_{P,i} = \{(m'', r_{P,neu}) | \exists r \in R : (m'', r) \in K_i \wedge ((m, r) \in K_r \vee (r, m') \in K_r) \wedge m, m', m'' \in M\}$  repräsentiert.

Den Kanten der Teilmenge  $K_{P,vi}$  wird der Typ  $vi$ , denen der Teilmenge  $K_{P,pi}$  der Typ  $pi$  und denen der Teilmenge  $K_{P,i}$  wird der Typ  $i$  zugeordnet. Weiterhin werden den Kanten der Teilmengen  $K_{P,vi}$  und  $K_{P,pi}$ , die mit dem Rückreaktionsknoten

verbunden sind  $((m', r_{P,neu})$  und  $(r_{P,neu}, m))$ , die Stöchiometrien der entsprechenden Ursprungskanten  $((r, m')$  und  $(m, r))$  zugeordnet.

- $A_P \subseteq A$  die Menge der Attribute in  $G_{Petri-Netz}$  entspricht einer Teilmenge der Attribute in  $G_{Abstrakt}$ ,  $A_P = \{Typ, Stöchiometrie, Lokalisation, Bezeichner, Menge, Kapazität, Rate\}$ , wobei die *Rate* nur eine stochastische Schaltrate oder die leere Menge  $\{\}$  (wenn keine stochastische Schalraten oder nur kinetische Ratengleichung vergeben wurden) ist.



**Abbildung 5.5:** Ableitung reversibler Reaktionen des *abstrakten Graphen* (links) in irreversible Hin- und Rückreaktionen des *Petri-Netz Graphen* (rechts): a) Reversible Reaktionen, b) Reversible Export- (oben) und Importreaktionen, c) Inhibition reversibler Reaktionen. Es werden neue Kantenarten gesetzt (blaue Färbung) und weitere irreversible Kanten ( $vi, pi$ ) und Reaktionsknoten ( $R'$ ) hinzugefügt (grüne Färbung).

### 5.2.2.2 Analysen

Die Analyse von Petri-Netzen erfolgt je nach Definition qualitativ oder quantitativ im zugehörigen metabolischen Modell (siehe Kapitel 3 Abschnitt 3.3 für Details).

Die qualitative Analyse basiert auf der Umformung des *Petri-Netz Graphen* in ein lineares Gleichungssystem. Die Lösung dessen resultiert in P- und T-Invarianten, die bestimmte Metabolite (*Plätze*) und Reaktionen (*Transitionen*) als invariant ermitteln. Des Weiteren kann der Erreichbarkeits- bzw. Überdeckungsgraph (bei unbeschränkten Petri-Netzen) berechnet werden. Die Knoten beschreiben die Zustände des Petri-Netzes (Metabolitkonzentrationen) und die Kanten den Zustandsübergang (Durchführung einer Reaktion).

Zur quantitativen Petri-Netz Analyse (stochastische Simulation) werden Zufallszahlen mit dem Gewicht der Reaktionen zu einer Schalthrscheinlichkeit verrechnet. Ein Schritt im Petri-Netz entspricht somit der Durchführung einer Reaktion mit der höchsten Wahrscheinlichkeit und der daraus resultierenden aktualisierten Metabolitkonzentrationen.

Für jeden Simulationsschritt wird dieser Prozess durchlaufen. Das Ergebnis der stochastischen Simulation sind Metabolitkonzentrationen und Reaktionsraten über die Anzahl der Simulationsschritte.

### 5.2.3 Stöchiometrisches Modell

Ein metabolisches Modell ohne Dynamik ist ein stöchiometrisches Modell. An dieser Stelle werden die regulatorischen Effekte durch Enzyme und Inhibitoren nicht betrachtet, weil für dieses Modell kinetische Informationen keine Verwendung finden (siehe Kapitel 3 Abschnitt 3.2).

#### 5.2.3.1 Graph

Ein gerichteter, attributierter, bipartiter Graph  $G_{Stöchiometrie}$  bildet dieses Modell wie folgt ab:  $G_{Stöchiometrie} = (M_S, R_S, K_S, A_S)$  besteht aus den beiden endlichen Knotenmengen Metabolit  $M_S$  und Reaktionen  $R_S$ , wobei beide Mengen disjunkt sind:  $M_S \cap R_S = \emptyset$ . Weiterhin enthält dieser Graph eine endliche gerichtete Kantenmenge  $K_S \subseteq (M_S \times R_S) \cup (R_S \times M_S)$  und eine endliche Menge von Attributen  $A_S = \{Typ, Stöchiometrie, Lokalisation, Bezeichner, Schranken, Zielfunktion\}$ . Der Graph  $G_{Stöchiometrie}$  kann wie folgt aus dem *abstrakten Graphen*  $G_{Abstrakt}$  abgeleitet werden:

- $M_S = M_{up}$  die Menge der Metabolite in  $G_{Stöchiometrie}$  entspricht einer Teilmenge der Metabolite in  $G_{Abstrakt}$ . Es werden nur die Metabolite in  $G_{Stöchiometrie}$  übertragen, die durch Verbrauchs- und Produktionskanten mit Reaktionen verbunden sind. Somit werden die mit Inhibitions-kanten verbundenen Metabolite, die in Reaktionen nicht zusätzlich verbraucht oder produziert werden, ausgeschlossen.
- $R_S = R$  die Menge der Reaktionen in  $G_{Stöchiometrie}$  entspricht der Menge der Reaktionen in  $G_{Abstrakt}$ .
- $K_S = K_{ir} \cup K_r$  die Menge der Kanten in  $G_{Stöchiometrie}$  entspricht einer Teilmenge der Kanten in  $G_{Abstrakt}$ . Es werden alle eingehenden und ausgehenden Kanten von Reaktionen abgeleitet, die nicht vom Typ Inhibition sind.
- $A_S \subseteq A$  die Menge der Attribute in  $G_{Stöchiometrie}$  entspricht einer Teilmenge der Attribute in  $G_{Abstrakt}$ ,  $A_S = \{Typ, Stöchiometrie, Lokalisation, Bezeichner, Schranken, Zielfunktion\}$ , siehe Abbildung 5.6.

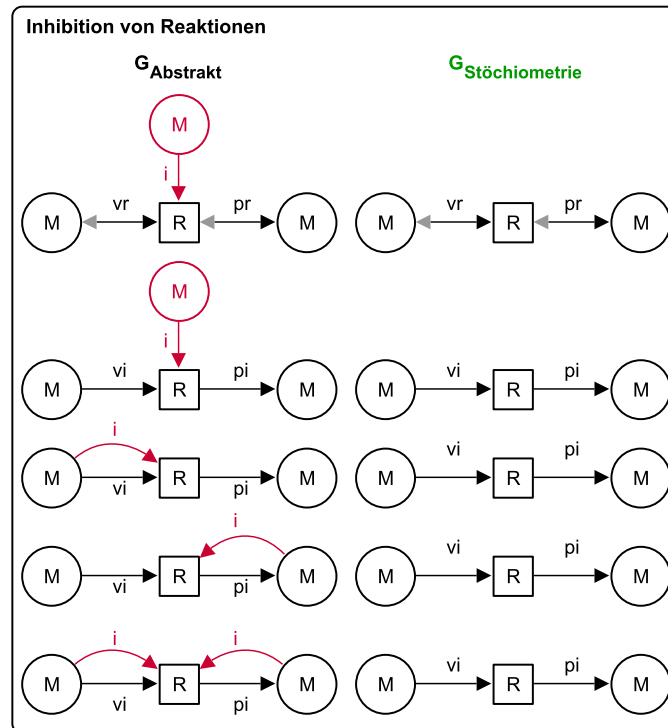
#### 5.2.3.2 Analysen

Die stöchiometrische Analyse untersucht das metabolische Modell zum einen strukturell mit der *Dead-End* Analyse und zum anderen mit optimierungsbasierten Methoden wie *Gap-Finding*, Flussbilanzanalyse (*FBA*), Flussvariabilitätsanalyse (*FVA*), Robustheitsanalyse (*RA*) und Knockout-Analyse (*KA*) (Details Kapitel 3 Abschnitt 3.4).

Zur strukturellen Analyse wird der *stöchiometrische Graph* in eine stöchiometrische Matrix umgeformt, die untersucht wird. Das Ergebnis dieser *Dead-End* Analyse besteht aus bestimmten Metaboliten die im Modell unverbunden sind.

Für die optimierungsbasierten Methoden wird der *stöchiometrische Graph* in ein lineares Gleichungssystem umgeformt, welches durch Optimierung gelöst wird. Die





**Abbildung 5.6:** Ableitung von inhibierten Reaktionen des *abstrakten Graphen* (links) in Reaktionen ohne Inhibition des *stöchiometrischen Graphen* (rechts). Inhibitions-kanten und die zugehörigen Metabolitknoten (rote Färbung) werden nicht abgeleitet, wenn diese nicht gleichzeitig mit anderen Kantentypen an Reaktionen teilnehmen.

*Gap-Finding* Analyse ermittelt bestimmte Metabolite funktionell: nicht produzierbare, nicht verbrauchbare und davon abhängige Metabolite. Die Ergebnisse der *FBA* entsprechen einer Flussverteilung (metabolitabhängige Flusswerte). Die *FVA* bestimmt den minimal und maximal möglichen Flusswert von Reaktionen. Resultierende Analyseergebnisse der *RA* sind Sensitivitätskurven und die Ergebnisse der *KA* Sensitivitätswerte.

## 5.2.4 Topologische Modelle

Ein metabolisches Modell kann rein strukturell, in Form aller bisher definierter Graphen, topologisch analysiert werden. Die am häufigsten verwendeten Formen zur topologischen Analyse sind die beiden unipartiten Graphen, *Metabolit-* und *Reaktionsgraph* (siehe Kapitel 3 Abschnitt 3.5), die in dieser Arbeit als topologische Modelle bezeichnet werden. Diese unipartiten Graphen werden durch die Projektion auf eine der beiden Knotenmengen von  $G_{\text{Abstrakt}}$  ( $M$  oder  $R$ ) erstellt.

### 5.2.4.1 Graphen

**Metabolitgraph** Der *Metabolitgraph*  $G_{\text{Metabolit}} = (M_M, K_M, A_M)$  ist ein gerichteter, attributierter, unipartiter Graph, bestehend aus einer nicht-leeren, endlichen Knotenmenge Metabolite und einer endlichen Menge gerichteter Kanten  $K_M \subseteq (M_M \times M_M)$ . Weiterhin enthält der Graph eine endliche Menge von Attributen  $A_M = \{\text{Lokalisation}, \text{Bezeichner}\}$ . Diese Mengen werden wie folgt vom *abstrakten Graphen*  $G_{\text{Abstrakt}}$  abgeleitet:

- $M_M = M_{vp}$  die Menge der Metabolite in  $G_{Metabolit}$  entspricht einer Teilmenge der Metabolite in  $G_{Abstrakt}$ . Es werden nur die Metabolite in  $G_{Metabolit}$  übertragen, die durch Verbrauchs- und Produktionskanten mit Reaktionen verbunden sind. Somit werden die mit Inhibitions-kanten verbundenen Metabolite, die in Reaktionen nicht zusätzlich auch verbraucht oder produziert werden, ausgeschlossen.
- $K_M = \{(m, m') | \exists r \in R : ((m, r) \in K_r \wedge (r, m') \in K_r) \vee ((m, r) \in K_{ir} \wedge (r, m') \in K_{ir})\} \wedge m, m' \in M \} \cup \{(m', m) | \exists r \in R : (m, r) \in K_r \wedge (r, m') \in K_r \wedge m, m' \in M\}$  die Menge der Kanten in  $G_{Metabolit}$  verbinden Metabolitknoten miteinander, die in Reaktionen des *abstrakten Graphen* verbraucht und produziert werden, siehe Abbildung 5.7.

Die Ableitung von z. B. reversiblen Reaktionen des  $G_{Abstrakt}$  ergibt antiparallele Kanten zwischen den Metabolitknoten, die zur Übersichtlichkeit (Reduktion von Kanten und Überschneidungen) durch Doppelpfeile dargestellt werden. Eine Hauptrichtung (vom Substrat zum Produkt wie im *abstrakten Graphen*) wird nicht angegeben. Diese Unterscheidung kann nach der Ableitung nicht nachvollzogen werden, wenn z. B. mehrere Reaktionen ein und dieselben Metabolite wechselseitig verbrauchen und produzieren.

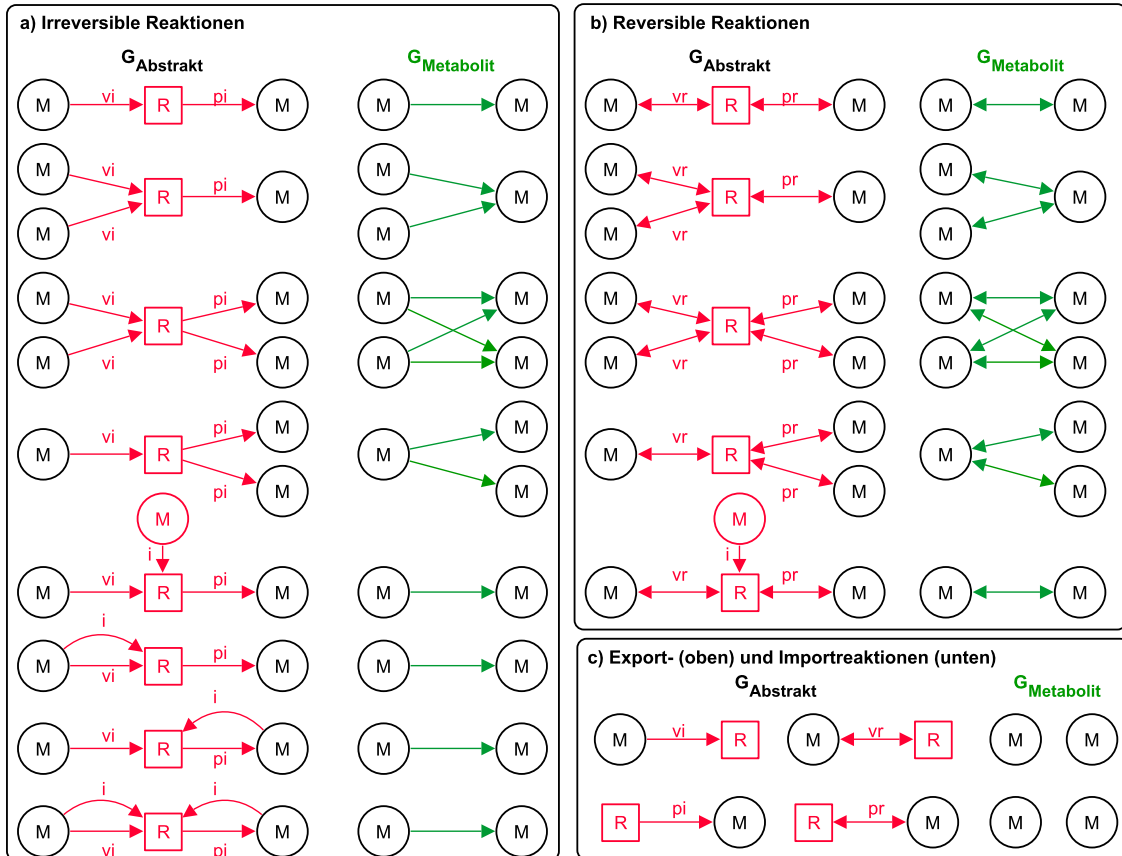
Existieren in  $G_{Abstrakt}$  Import- oder Exportreaktionen, deren Metabolite nicht mit weiteren Reaktionen verbunden sind, resultiert die Ableitung des  $G_{Metabolit}$  in unverbundenen Metabolitknoten.

- $A_M \subseteq A$  ist die Menge der Attribute in  $G_{Metabolit}$  und entspricht einer Teilmenge der Attribute in  $G_{Abstrakt}$ ,  $A_M = \{Typ_{neu}, Lokalisation, Bezeichner\}$ .

**Reaktionsgraph** Der *Reaktionsgraph*  $G_{Reaktion} = (R_R, K_R, A_R)$  ist ein gerichteter, attributierter, unipartiter Graph, bestehend aus einer nicht-leeren, endlichen Knotenmenge, Reaktionen  $R_R$  und einer endlichen Menge gerichteter Kanten  $K_R \subseteq (R_R \times R_R)$ . Weiterhin enthält der Graph eine endliche Menge von Attributen  $A_R = \{Bezeichner\}$ . Der *Reaktionsgraph* wird wie folgt vom *abstrakten Graphen*  $G_{Abstrakt}$  abgeleitet:

- $R_R = R$  entspricht der Menge der Reaktionen in  $G_{Abstrakt}$ .
- $K_R$  die Menge der Kanten in  $G_{Reaktion}$  verbinden Reaktionsknoten, die über einen gemeinsamen Metaboliten miteinander verbunden sind, siehe Abbildung 5.8. Die Kantenmenge  $K_R = K_{R,a} \cup K_{R,b} \cup K_{R,c}$  wird aus drei Teilmengen gebildet, die aufgrund der Betrachtung von jeweils zwei aufeinanderfolgenden Reaktionen und deren gemeinsamen Metaboliten abgeleitet werden. Die Teilmenge  $K_{R,a}$  bildet Kanten, die im *abstrakten Graphen* von einem Reaktionsknoten  $r$  zu einem Metabolitknoten  $s$  zu einem anderen Reaktionsknoten  $r'$  verlaufen:  $K_{R,a} = \{(r, r') | \exists m \in M : (r, m) \in K_{ir} \wedge Typ((r, m)) = pi \wedge (m, r') \in K_{ir} \wedge Typ((m, r')) = vi \wedge r, r' \in R\}$ , siehe Abbildung 5.8 a) oben.

Die Teilmenge  $K_{R,b}$  bildet antiparallele Kanten, die entstehen wenn zwei Reaktionsknoten im *abstrakten Graphen* über einen gemeinsamen Metabolitknoten entweder durch zwei reversible Kanten oder einer reversiblen und einer irreversiblen Kante verbunden sind:



**Abbildung 5.7:** Ableitung des *abstrakten Graphen* (links) in den *Metabolitgraphen* (rechts): a) Irreversible Reaktionen, b) Reversible Reaktionen und c) Export- (oben) und Importreaktionen (unten). Reaktionsknoten, Kanten und Metabolitknoten, die mit einer Inhibitionskante mit der Reaktion verbunden sind (rote Färbung), werden nicht abgeleitet. Es werden gerichtete Kanten zwischen den verbleibenden Metabolitknoten hinzugefügt (grüne Färbung). Antiparallele Kanten werden durch Doppelpfeile gekennzeichnet.

$$K_{R,b} = \{(r, r') | \exists m \in M : (((r, m) \in K_{ir} \wedge (m, r') \in K_r) \vee ((r, m) \in K_r \wedge (m, r') \in K_{ir}) \vee ((r, m) \in K_r \wedge (m, r') \in K_r)) \wedge r, r' \in R\} \cup \{(r', r) | \exists m \in M : (((r, m) \in K_{ir} \wedge (m, r') \in K_r) \vee ((r, m) \in K_r \wedge (m, r') \in K_{ir}) \vee ((r, m) \in K_r \wedge (m, r') \in K_r)) \wedge r, r' \in R\},$$

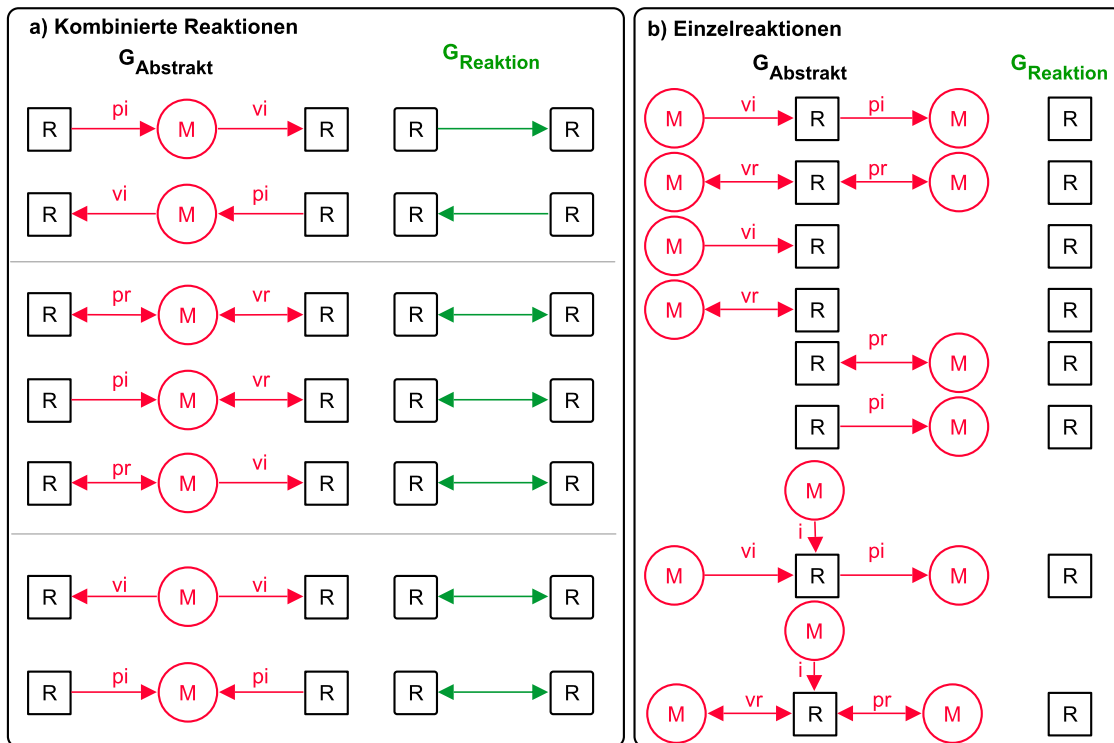
siehe Abbildung 5.8 a) Mitte.

Die Teilmenge  $K_{R,c}$  bildet antiparallele Kanten, die entstehen wenn zwei Reaktionsknoten im *abstrakten Graphen* über einen gemeinsamen Metabolitknoten, entweder durch zwei irreversible Verbrauchs- oder zwei irreversible Produktionskanten verbunden sind:  $K_{R,c} = \{(r, r') | \exists m \in M : (((r, m) \in K_{ir} \wedge Typ((r, m)) = vi \wedge (m, r') \in K_{ir} \wedge Typ((m, r')) = vi) \vee ((r, m) \in K_{ir} \wedge Typ((r, m)) = pi \wedge (m, r') \in K_{ir} \wedge Typ((m, r')) = pi)) \wedge r, r' \in R\} \cup \{(r', r) | \exists m \in M : (((r, m) \in K_{ir} \wedge Typ((r, m)) = vi \wedge (m, r') \in K_{ir} \wedge Typ((m, r')) = vi) \vee ((r, m) \in K_{ir} \wedge Typ((r, m)) = pi \wedge (m, r') \in K_{ir} \wedge Typ((m, r')) = pi)) \wedge r, r' \in R\}$ , siehe Abbildung 5.8 a) unten.

Die antiparallelen Kanten der Teilmengen  $K_{R,b}$  und  $K_{R,c}$  werden zur Übersichtlichkeit (Reduktion von Kanten und Überschneidungen) durch Doppelpfeile dargestellt.

Existieren in  $G_{\text{Abstrakt}}$  Reaktionen, deren Metabolite nicht mit weiteren Reaktionen verbunden sind, resultiert die Ableitung des  $G_{\text{Reaktion}}$  in unverbundenen Reaktionsknoten, siehe Abbildung 5.8 b).

- $A_R \subseteq A$  ist die Menge der Attribute in  $G_{Reaktion}$  und entspricht einer Teilmenge der Attribute in  $G_{Abstrakt}$ ,  $A_M = \{Bezeichner\}$ .



**Abbildung 5.8:** Ableitung des *abstrakten Graphen* (links) in den *Reaktionsgraphen* (rechts): a) Kombinierte Reaktionen und b) Einzelreaktionen. Metabolitknoten und Kanten (rote Färbung) werden nicht abgeleitet. Es werden gerichtete Kanten zwischen den verbleibenden Reaktionsknoten hinzugefügt (grüne Färbung). Antiparallele Kanten werden durch Doppelpfeile gekennzeichnet.

### 5.2.4.2 Analysen

Metabolische Modelle werden topologisch durch die Umformung des *Metabolitgraphen* oder *Reaktionsgraphen* in eine Adjazenzmatrix untersucht. Die topologische Analyse der kürzesten Wege zwischen zwei Knoten resultiert in Teilgraphen oder den untersuchten Graphen selbst. Die Zentralitätsanalyse mit den verschiedenen Zentralitätsmaßen resultiert in einer geordneten Liste von Zentralitätswerten. Eine detaillierte Erläuterung der topologischen Analysen befindet sich in Kapitel 3 Abschnitt 3.5.

### 5.2.5 Zusammenfassung der Analyseergebnisse

Die bereits im vorherigen Abschnitt vorgestellten Analysen der verschiedenen Modellierungsformalismen stellen einen Ausschnitt vieler möglicher Analysemethoden dar. Weiterführende Analysemethoden verschiedener Modellierungsformalismen in der Systembiologie werden von Machado und Koautoren [106] vorgestellt. In dieser Arbeit liegt der Fokus auf typische Analysemethoden metabolischer Modellierungsformalismen. Deren zugehörige Analyseergebnisse werden im Folgenden zusammengetragen, in Datentypen kategorisiert und den Elementen des *abstrakten Graphen* zugeordnet, siehe Tabelle 5.1. Die erhobenen Analyseergebnisse können in zwei Datentypkategorien eingeteilt werden:

**Tabelle 5.1:** Zusammenfassung typischer Analysemethoden und entsprechende Analyseergebnisse, die in Datentypen kategorisiert und den Graphenelementen (Metabolitknoten ( $M$ ), Reaktionsknoten ( $R$ ), Kanten ( $K$ )) des *abstrakten Graphen* zugeordnet werden. Fußnoten: <sub>1</sub> Analyseergebnisse die Hin- und Rückreaktionsknoten im Petri-Netz betreffen, werden den entsprechenden reversiblen Reaktionsknoten des *abstrakten Graphen* zugeordnet, <sub>2</sub> Analyseergebnisse zu Kanten im *Metabolit- und Reaktionsgraphen* werden verschiedenen Kanten und Knoten des *abstrakten Graphen* zugeordnet.

Modellierungsformalismus	Typische Analysemethoden		Analyseergebnisse	Datentypen	GAbstrakt		
					M	R	K
Kinetische Modellierung	Kinetische Analyse		Metabolitkonzentration und Reaktionsraten über die Zeit	Vektor zeitabhängiger, numerischer Werte	x	x	
Petri-Netz Modellierung	Invariantenanalyse		P- und T-Invarianten	Vektor numerischer Werte	x <sub>1</sub>	x <sub>1</sub>	
	Erreichbarkeitsanalyse		Erreichbarkeitsgraph/ Überdeckungsgraph	Graph	x <sub>1</sub>	x <sub>1</sub>	
	Quantitative Analyse (stochastisch)		Metabolitkonzentration und Reaktionsraten über Schritte	Vektor schrittabhängiger numerischer Werte	x <sub>1</sub>	x <sub>1</sub>	
Stöchiometrische Modellierung	Stöchiometrische Analyse		Dead-Ends	Knoten	x		
	Optimierungs-basierte Analyse	Gap-Finding	Gaps	Knoten	x		
		FBA	Flussverteilung	Vektor numerischer Werte			x
		RA	Sensitivitätskurve	Vektor flusswert-abhängiger numerischer Werte		x	
		KA	Sensitivitätswert	Vektor numerischer Werte		x	
		FVA	Min/Max Flusswerte von Reaktionen	Vektor numerischer Wertepaare		x	
Topologische Modellierung	Zentralitätsanalyse		Zentralitätswerte	Vektor numerischer Werte	x	x	
	Analyse kürzester Wege		Kürzester Weg	Graph	x <sub>2</sub>	x <sub>2</sub>	x <sub>2</sub>

Vektoren numerischer Werte und Graphenelemente. Beide Datentypkategorien werden zur modellintegrierten Visualisierung entsprechenden Graphenelementen (Metabolitknoten ( $M$ ), Reaktionsknoten ( $R$ ) und Kanten ( $K$ )) des *abstrakten Graphen* zugeordnet:

- Analyseergebnisse bezüglich Metaboliten werden Metabolitknoten zugeordnet:
  - Metabolitkonzentrationen über die Zeit und Schritte,
  - P-Invarianten,
  - *Gaps* und *Dead-Ends*,
  - Zentralitätswerte und kürzeste Wege (bezüglich der Analyse des *Metabolitgraphen*)
- Analyseergebnisse bezüglich Reaktionen werden Reaktionsknoten zugeordnet:
  - Reaktionsraten über die Zeit und Schritte,
  - T-Invarianten,
  - Sensitivitätskurven,
  - Sensitivitätswerte,
  - Min/Max Flusswerte,
  - Zentralitätswerte und kürzeste Wege (bezüglich der Analyse des *Reaktionsgraphen*).

3. Analyseergebnisse bezüglich der Kombination von Metaboliten und Reaktionen werden den Kanten zugeordnet:

- Flussverteilung

Die Formalisierung dieser Datenzuordnung und deren Visualisierung wird im nächsten Abschnitt erläutert.

## 5.3 Integration und Visualisierung der Analyseergebnisse

Ein metabolisches Modell repräsentiert als *abstrakter Graph* wird durch die Ableitungen in *spezielle Graphen* mit verschiedenen Modellierungsformalimen analysiert. In dieser Arbeit liegt der Fokus auf typischen Analysemethoden metabolischer Modellierungsformalimen, die strukturelle und dynamische Eigenschaften des Modells erfassen (weiterführende Analysemethoden in der Systembiologie werden in [106] vorgestellt). Diese Analyseergebnisse werden im Folgenden mittels Datenzuordnungsfunktionen in die einheitliche Visualisierung des *abstrakten Graphen* integriert.

### 5.3.1 Datenzuordnungsfunktionen

Die Modellierungsformalimen verwenden *spezielle Graphen* zur Analyse. Deshalb bedarf es der Definition von Datenzuordnungsfunktionen, die die Integration der Analyseergebnisse in Graphenelemente (Metabolitknoten ( $M$ ), Reaktionsknoten ( $R$ ), Kanten ( $K$ )) des *abstrakten Graphen* vornehmen.

#### 5.3.1.1 Datenzuordnungsfunktion für numerische Werte

Die numerischen Werte der Vektoren ( $nv \in NV$ ) werden mit der Datenzuordnungsfunktion  $zn$  den Graphenelementen (Metabolitknoten ( $M$ ), Reaktionsknoten ( $R$ ) und Kanten ( $K$ )) zugeordnet  $zn : M, R, K \rightarrow NV$ . Ein Vektor kann aus numerischen Werten (z. B. P-Invariante), Wertepaaren (z. B. Min/Max Flusswerte von Reaktionen) und zeit-, schritt- und flusswertabhängiger numerischer Werte (z. B. Metabolitkonzentrationen über die Zeit oder Schritte und Sensitivitätskurven) zusammengesetzt sein.

#### 5.3.1.2 Datenzuordnungsfunktion für Graphenelemente

Einige Analysen resultieren in Graphenelementen, die dem *abstrakten Graphen* zugeordnet werden. Die Erreichbarkeitsanalyse im Petri-Netz Modellierungsformalismus resultiert in einem Erreichbarkeits-/Überdeckungsgraphen (kurz EG oder ÜG), deren Knoten alle erreichbaren Zustände eines Petri-Netzes, ausgehend von einer Anfangsmarkierung, darstellen. Ein Zustand ist die Markierung des Petri-Netzes, die nach einer Schaltabfolge von Transitionen erreicht wird. Die Kanten zwischen den Zuständen im EG oder ÜG entsprechen den Transitionen des Petri-Netzes, die schalten, um von einem Zustand in den nächsten zu gelangen. Übertragen auf das metabolische Modell entspricht ein Zustand der aktuellen Konzentrationen aller Metabolite, die nach Durchführung von biochemischen Reaktionen entstehen. Die Integration der Metabolitkonzentrationen (beschrieben durch Knoten des EG oder ÜG), welche numerischen Werten entsprechen, erfolgt über die Verwendung der Datenzuordnungsfunktion  $zn$  (Abschnitt 5.3.1), wobei den entsprechenden Metabolitknoten ( $M$ ) im *abstrakten Graphen* die zugehörige Konzentration zugeordnet wird.

Die Integration von Graphelmenten der *speziellen Graphen* auf Graphelemente des *abstrakten Graphen* erfolgt über die Datenzuordnungsfunktion  $zg : M, R, K \rightarrow M_x, R_x, K_x$ , wobei  $x$  Platzhalter für  $P$  Petri-Netz,  $S$  Stöchiometrie,  $K$  Kinetik,  $M$  Metabolit oder  $R$  Reaktion ist.

Diese Datenzuordnungsfunktion wird bei der Integration der Ergebnisse der Erreichbarkeitsanalyse verwendet, um Reaktionsknoten des *Petri-Netz Graphen* (beschrieben durch Kanten des EG oder ÜG) in Reaktionsknoten des *abstrakten Graphen* zu integrieren:  $zg : R \rightarrow R_P$ . Zu beachten ist, dass aufgrund der Aufspaltung von reversiblen Reaktionsknoten vom *abstrakten Graphen*, in Hin- und Rückreaktionsknoten des *Petri-Netz Graphen* beide Knoten in den dazugehörigen Reaktionsknoten in  $G_{Abstrakt}$  integriert werden.

Die stöchiometrische und *Gap-Finding* Analysen der stöchiometrischen Modellierung resultieren in einer Menge von Metabolitknoten. Die Menge der Metabolitknoten des *stöchiometrischen Graphen* ist eine Teilmenge der Metabolitknotenmenge im *abstrakten Graphen*. An dieser Stelle wird die Datenzuordnungsfunktion  $zg : M \rightarrow M_S$  verwendet, um die Metabolitknoten des *stöchiometrischen Graphen* denen des *abstrakten Graphen* zuzuordnen.

Die Analyse des kürzesten Weges der topologischen Modellierung generiert kürzeste Wege (Teilgraphen oder die *speziellen Graphen* (*Metabolit-* oder *Reaktionsgraph*) selbst). Die resultierende Menge Metabolit- oder Reaktionsknoten ist eine Teilmenge der Knoten im *abstrakten Graphen*. Die Zuordnung dieser Knoten von den *speziellen Graphen*, in den *abstrakten Graphen* erfolgt unter Verwendung der Datenzuordnungsfunktion  $zg$ , die im Fall des *Metabolitgraphen* in Form von  $zg : M \rightarrow M_M$  und im Fall des *Reaktionsgraphen* in Form von  $zg : R \rightarrow R_R$  definiert ist.

Die Menge der resultierenden Kanten der topologischen Analyse ist verschieden von den Kanten im *abstrakten Graphen* und wird entsprechend der folgenden Vorschrift zugeordnet: Eine Kante, die Analyseergebnis vom *Metabolitgraphen* (*Reaktionsgraphen*) ist und sich zwischen zwei Metabolitknoten (Reaktionsknoten) befindet, wird allen Kanten im *abstrakten Graphen* zugeordnet, die sich zwischen denselben beiden Metabolitknoten (Reaktionsknoten) befinden. Demzufolge wird eine Kante des *Metabolit-* oder *Reaktionsgraphen* mehreren Kanten des *abstrakten Graphen* zugeordnet, mindestens einer (ir-)reversiblen Verbrauchskante und einer (ir-)reversiblen Produktkante. Die Zuordnung der Kanten erfolgt durch die Datenzuordnungsfunktion:  $zg : K \rightarrow K_M$  ( $zg : K \rightarrow K_R$ ), sodass eine Kante des *Metabolitgraphen* (*Reaktionsgraphen*)  $k \in K_M$  ( $k \in K_R$ ) mehreren Kanten  $K$  des *abstrakten Graphen* zugeordnet werden kann.

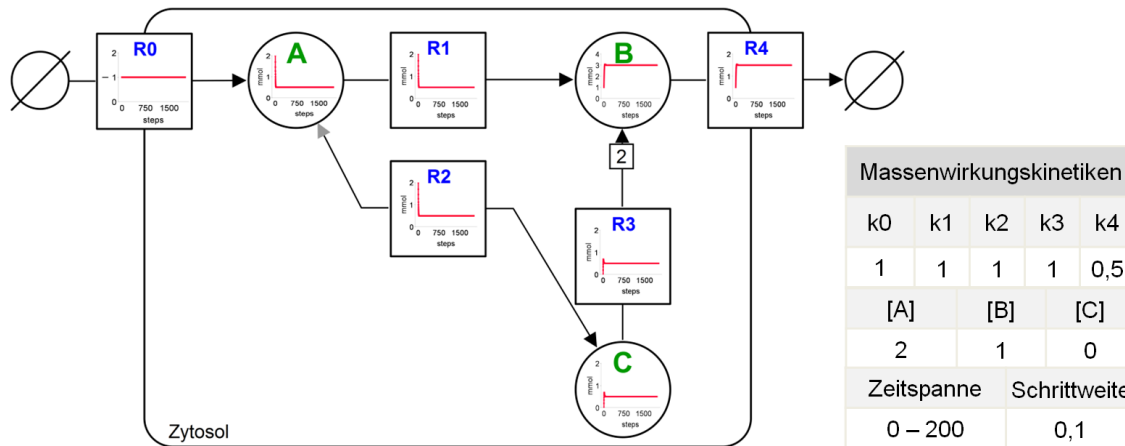
Nachdem die vielfältigen Analyseergebnisse den entsprechenden Graphelementen des *abstrakten Graphen* zugeordnet sind, erfolgt eine anschließende modellintegrierte Visualisierung dieser Resultate.

### 5.3.2 Visualisierung

Eine interaktive und dynamische Visualisierung der verschiedenen Analyseergebnisse erfolgt in der einheitlichen und standardisierten Repräsentation des *abstrakten Graphen*. Geeignete Visualisierungen ermöglichen eine intuitive und einfache Interpretation der Analyseergebnisse und Interaktionstechniken erlauben die Exploration struktureller und dynamischer Eigenschaften metabolischer Modelle. Anhand des in Kapitel 2 und Kapitel 3 vorgestellten metabolischen Beispiel Modells werden im Folgenden die Visualisierungen der Ergebnisse unterschiedlicher Modellierungsformalismen zum einen unabhängig voneinander und zum anderen integriert dargestellt.

### 5.3.2.1 Kinetische Modellierung

Die kinetische Analyse ermittelt für jeden Metaboliten und jede Reaktion des metabolischen Modells zeitabhängige, numerische Werte: die Metabolitkonzentrationen und Reaktionsraten. Die Analyseergebnisse werden in einem Diagramm zusammengefasst, welches die Metabolitkonzentrationen und Reaktionsraten über die Zeit veranschaulicht. Eine zusätzliche Möglichkeit der modellintegrierten Visualisierung wird durch die Integration der einzelnen Kurvenverläufe in die entsprechenden Knoten des *abstrakten Graphen* realisiert. Ein Metabolitknoten erhält somit die Konzentrationsverteilung der Metabolite über die Zeit und ein Reaktionsknoten die entsprechende Reaktionsrate über die Zeit, siehe Abbildung 5.9.



**Abbildung 5.9:** Visualisierung der Analyseergebnisse der kinetischen Analyse am Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Import-  $R_0$ , einer Exportreaktion  $R_4$ , zwei irreversiblen  $R_1, R_3$  und einer reversiblen Reaktion  $R_2$  sowie drei Metaboliten  $A, B, C$ . Reversible Reaktionen werden durch Doppelpfeile dargestellt, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist. Integrierte Diagramme der Metabolitkonzentrationen und Reaktionsraten über die Zeit sind das Ergebnis der kinetischen Analyse (links). Die verwendeten Parameter (rechts): Massenwirkungskinetiken mit den kinetischen Ratenkonstanten für die Reaktionen  $R_0 - R_4$  ( $k_0 - k_4$ ), den Metabolitkonzentrationen  $[A], [B], [C]$ , der Zeitspanne  $0 - 200$  s und der Schrittweite  $0,1$  s der Simulation.

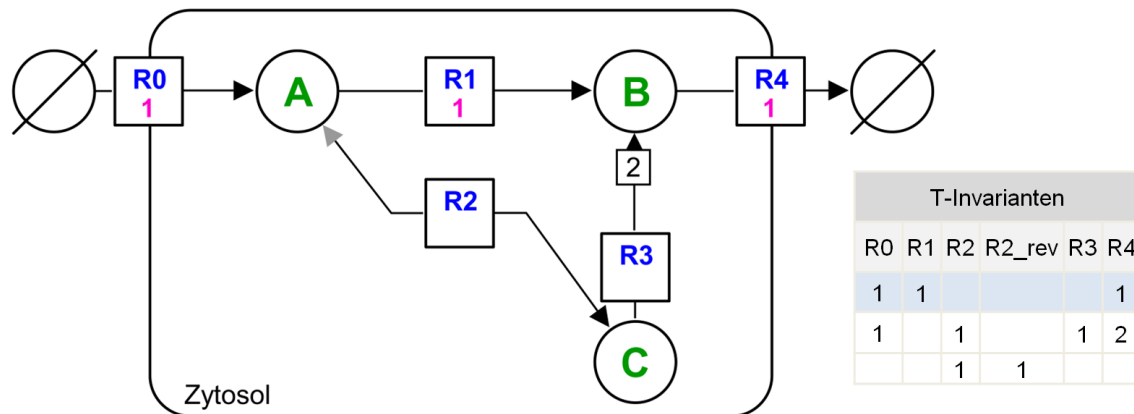
### 5.3.2.2 Petri-Netz Modellierung

Die Petri-Netz Analyse eines metabolischen Modells resultiert in schrittabhängigen, numerischen Werten, den Metabolitkonzentrationen und Reaktionsraten. Ebenso wie die Diagramme der kinetischen Analyse (Abschnitt 5.3.2.1) werden die Ergebnisse der Petri-Netz Analyse als zusammenfassendes Diagramm und in Form einer modellintegrierten Visualisierung (einzelne Diagramme pro Knoten) dargestellt (nicht gezeigt).

Die qualitative Petri-Netz Analyse besteht aus der Invarianten- und der Erreichbarkeitsanalyse. Die Invariantenanalyse erhebt P- und T-Invarianten, die in Tabellen für die jeweilige Menge der P- und T-Invarianten zusammengefasst werden. Die Integration der verschiedenen P- bzw. T-Invarianten in die Knoten des *abstrakten Graphen* erfolgt durch die Interaktionstechnik *Brushing & Linking* (siehe Kapitel 2 Abschnitt 2.3.3). Das Linking verknüpft die Einträge (Zeilen) der Invariantentabelle (je Zeile eine Invariante) mit der visuellen Darstellung des *abstrakten Graphen*. Die Auswahl einer Invariante führt zur Visualisierung dieser im *abstrakten Graphen* durch Brushing. In Abbildung 5.10 ist die



T-Invariantentabelle mit selektierter T-Invariante dargestellt, die im *abstrakten Graphen* durch die Integration der numerischen Werte in den Knoten visualisiert wird.



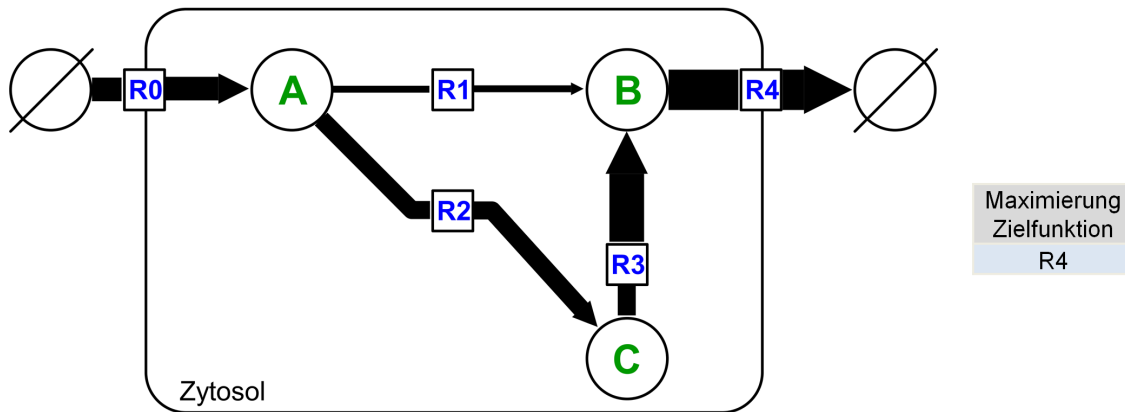
**Abbildung 5.10:** Visualisierung der Analyseergebnisse der Petri-Netz Invariantenanalyse am Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Importreaktion  $R_0$ , einer Exportreaktion  $R_4$ , zwei irreversiblen  $R_1, R_3$  und einer reversiblen Reaktion  $R_2$  sowie drei Metaboliten  $A, B, C$ . Reversible Reaktionen werden durch Doppelpfeile dargestellt, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist. Die Invariantenanalyse resultiert in drei T-Invarianten (rechts), wobei die erste T-Invariante in die Reaktionsknoten integriert visualisiert ist (links, Schaltrate in Magenta unten in den Reaktionsknoten).

Die Erreichbarkeitsanalyse resultiert in der Erstellung eines Erreichbarkeits- oder Überdeckungsgraphen (im Fall eines unbeschränkten metabolischen Modells). Diese Graphen repräsentieren Zustände (Knoten) des metabolischen Modells nach einer Abfolge von Reaktionen (Kanten). Ein Zustand enthält Metabolitkonzentrationen, die durch das Ausführen einer Reaktionen in einen anderen Zustand mit veränderten Metabolitkonzentrationen überführt wird. Die Interaktionstechnik *Brushing & Linking* verknüpft hier die Knoten und Kanten des Erreichbarkeits- oder Überdeckungsgraphen mit den Metabolit- und Reaktionsknoten des *abstrakten Graphen*. Wird im Erreichbarkeits- oder Überdeckungsgraphen ein Knoten ausgewählt, werden die entsprechenden Metabolitkonzentrationen in den Metabolitknoten des *abstrakten Graphen* visualisiert. Die Auswahl einer Kante im Erreichbarkeits- oder Überdeckungsgraphen resultiert in der Visualisierung eines Reaktionsknotens im *abstrakten Graphen*.

### 5.3.2.3 Stöchiometrische Modellierung

Die stöchiometrische Analyse eines metabolischen Modells erfolgt zum einen strukturell und zum anderen optimierungsbasiert. Die stöchiometrische *Dead-End* Analyse, also auch die optimierungsbasierte *Gap-Finding* Analyse, resultiert in Metabolitknoten. Diese Metabolitknoten werden im *abstrakten Graphen* entsprechend deren Kategorien visualisiert. Die *Dead-End* Metabolite und die davon abhängigen Metabolite, die nicht produziert oder verbraucht werden, erhalten eine bestimmte Färbung.

Die folgenden Ausführungen zur Visualisierung sind der Dissertation von Grafahrend-Belau entnommen und können in [60] nachgelesen werden. Die Flussverteilung der Flussbilanzanalyse (FBA) entspricht der Kombination von Metaboliten und Reaktionen. Zur Visualisierung der numerischen Flusswerte werden die Kantenbreiten im *abstrakten Graphen* skaliert, siehe Abbildung 5.11.



**Abbildung 5.11:** Visualisierung der Analyseergebnisse der optimierungsbasierten Analyse (*FBA*) am Beispiel eines kleinen metabolischen Modells in *SBGN-PD*. Das Modell besteht aus einer Importreaktion *R0*, einer Exportreaktion *R4*, zwei irreversiblen *R1*, *R3* und einer reversiblen Reaktion *R2* sowie drei Metaboliten *A*, *B*, *C*. Unter Maximierung der Zielfunktion *R4* (rechts, *qFBA*) wird die *FBA* durchgeführt. Die Skalierung der Kantenbreiten visualisiert die Flussverteilung integriert in das Modell (links).

Zusätzlich können Parameter schrittweise variiert werden, sodass die davon abhängigen Flussverteilungen entsprechend der gewählten Schrittweite dynamisch angepasst werden. Somit entsteht eine Animation.

Die weiteren optimierungsbasierten Analysen resultieren in Ergebnissen, die in den Reaktionsknoten des *abstrakten Graphen*, integriert visualisiert werden. Die Sensitivitätskurven als Resultat der Robustheitsanalyse (*RA*) werden in Form von Kurvendiagrammen, die Sensitivitätswerte (Resultat der Knockout-Analyse, *KO*) und die minimalen und maximalen Flusswerte (Resultat Flussvariabilitätsanalyse, *FVA*) werden als Balkendiagramme in den Reaktionsknoten visualisiert.

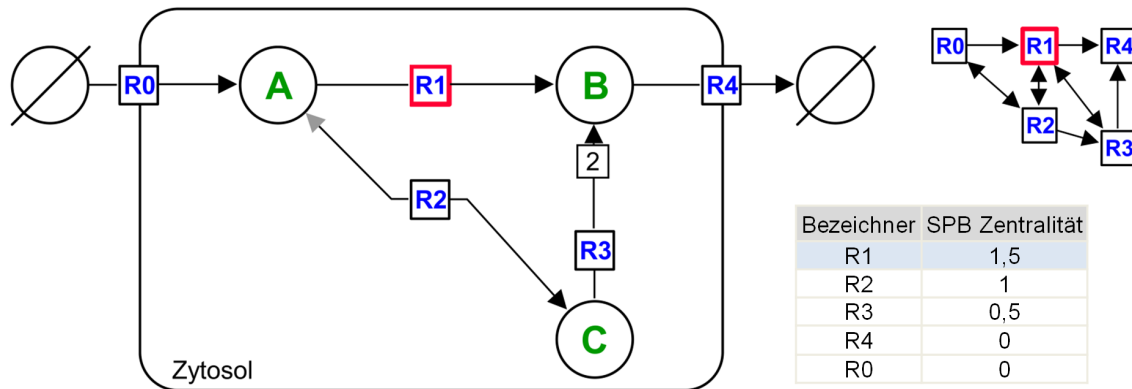
#### 5.3.2.4 Topologische Modellierung

Die topologische Analyse kürzester Weg eines metabolischen Modells resultiert in Graphen, deren Knoten und Kanten mittels einer Auswahl der entsprechenden Graphenelemente des *abstrakten Graphen* visualisiert werden.

Die verschiedenen Zentralitätsanalysen generieren geordnete Listen von Zentralitätswerten. Die Interaktionstechnik *Brushing & Linking* verknüpft die Zentralitätswerte mit den entsprechenden Knoten des *abstrakten Graphen*. Wird ein Wert oder Knoten ausgewählt, erfolgt die Visualisierung, siehe Abbildung 5.12.

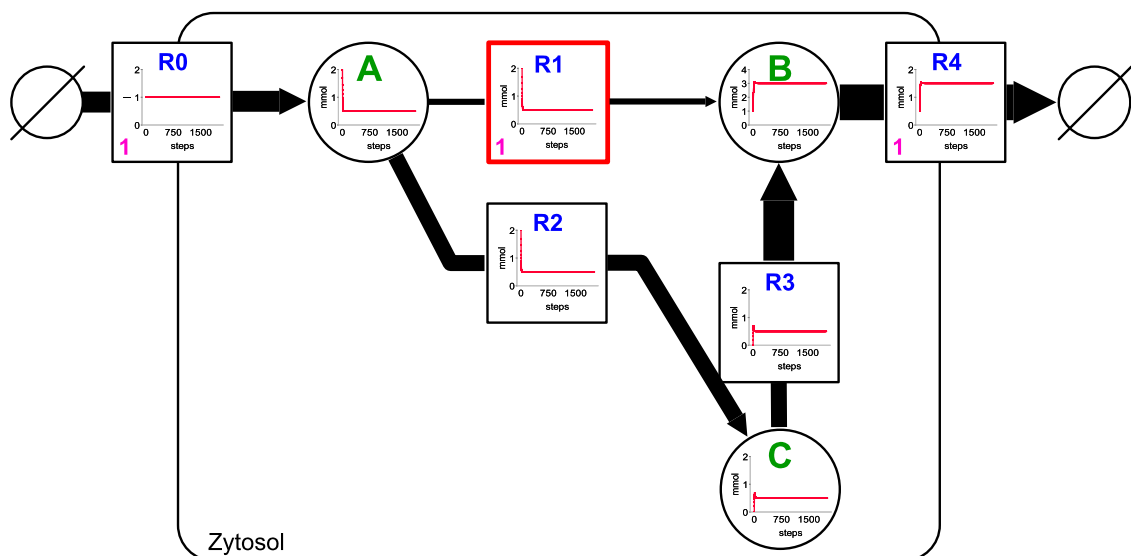
**Integration verschiedener Analyseergebnisse** Neben der Visualisierung der unterschiedlichen Analyseergebnisse nacheinander, ist auch eine integrierte Visualisierung möglich, siehe Abbildung 5.13.

Diese Integration ermöglicht die strukturellen und dynamischen Eigenschaften eines metabolischen Modells resultierend aus verschiedenen Formalismen im Kontext des Modells visuell zu untersuchen. Besonders im Fall unzureichender Experimentaldaten zur detaillierten Beschreibung der Reaktionsmechanismen jeder Reaktion des metabolischen Modells ist diese Integration von Vorteil. Ist z. B. ein Stoffwechselweg im Modell mit kinetischen Reaktionsmechanismen und Parametern versehen und andere Stoffwechselwege nicht, kann die kinetische Analyse für diesen Stoffwechselweg vorgenommen werden.



**Abbildung 5.12:** Visualisierung der Analyseergebnisse der topologischen Zentralitätsanalyse (*Shortest-Path Betweenness*, SPB Zentralität) am Beispiel eines kleinen metabolischen Modells in SBGN-PD. Das Modell besteht aus einer Importreaktion  $R_0$ , einer Exportreaktion  $R_4$ , zwei irreversiblen  $R_1$ ,  $R_3$  und einer reversiblen Reaktion  $R_2$ , sowie drei Metaboliten  $A$ ,  $B$ ,  $C$ . Der Knoten mit dem größten SPB Zentralitätswert ( $R_1$ ) ist in der Ergebnistabelle ausgewählt (rechts unten) und wird im Modell und im entsprechenden Reaktionsgraphen, der Grundlage dieser Analyse ist, mit entsprechender Auswahl dargestellt (rechts oben).

Die Ergebnisse werden dann in das Modell integriert. Damit die nicht detailliert, kinetisch beschriebenen Stoffwechselwege dennoch quantitativ analysiert werden können, wird z. B. die Petri-Netz Analyse durchgeführt. Diese bedarf der Angabe von stochastischen Schaltraten und kann die fehlenden kinetischen Ergebnisse approximieren.

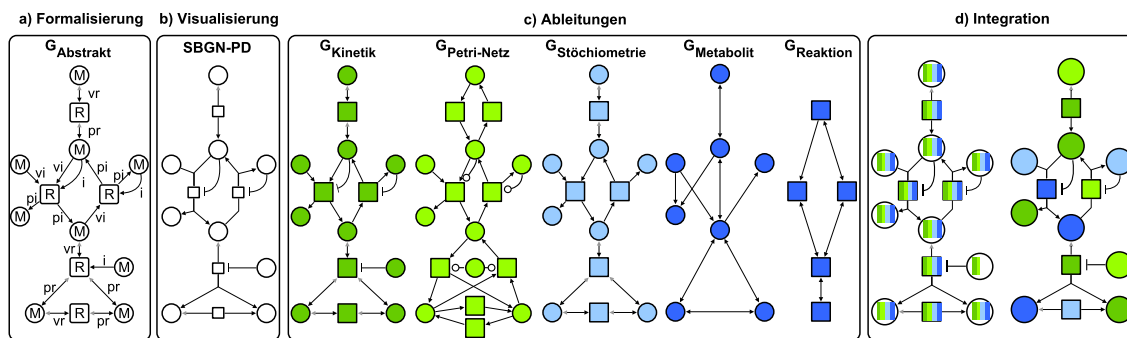


**Abbildung 5.13:** Integrierte Visualisierung der Analyseergebnisse am Beispiel eines kleinen metabolischen Modells in SBGN-PD. Das Modell besteht aus einer Importreaktion  $R_0$ , einer Exportreaktion  $R_4$ , zwei irreversiblen  $R_1$ ,  $R_3$  und einer reversiblen Reaktion  $R_2$  sowie drei Metaboliten  $A$ ,  $B$ ,  $C$ . Reversible Reaktionen werden durch Doppelpfeile dargestellt, wobei die Hauptrichtung (vom Substrat zum Produkt) durch die schwarze Pfeilspitze gekennzeichnet ist. Diagramme in den Metabolit- und Reaktionsknoten resultieren aus der kinetischen Analyse. Die Schaltrate in den Reaktionsknoten (linke untere Ecke in pink) zeigt eine T-Invariante als Resultat der Petri-Netz Invariantenanalyse. Skalierter Kantenbreiten stellen das Ergebnis der qFBA unter Maximierung der Zielfunktion  $R_4$  dar. Der größte SPB Zentralitätswert aus der topologischen Analyse des Reaktionsgraphen wird durch Markierung des Reaktionsknoten  $R_1$  visualisiert.

Ebenso kann die Flussbilanzanalyse unter Angabe einer Zielfunktion verwendet werden, um durch die berechnete Flussverteilung im Modell das Verhalten des modellierten biologischen Systems abzuleiten. Weitere strukturelle Eigenschaften können durch die topologische Analyse (z. B. Zentralitätsanalyse) hervorgebracht werden. Weiterhin ist es empfehlenswert ein bereits detailliert quantitativ beschriebenes metabolisches Modell, unter Verwendung zusätzlicher Modellierungsformalismen, aus weiteren Perspektiven zu analysieren. Die sich ergänzenden Analyseergebnisse und auch die Integration von Analyseergebnissen in anderen Formalismen zur weiterführenden Analyse unterstützen die Interpretation der strukturellen und dynamischen Eigenschaften des Modells.

## 5.4 Zusammenfassung

In der erarbeiteten Methodik, siehe Abbildung 5.14 konnte gezeigt werden, dass qualitative und quantitative metabolische Modelle in einer einheitlichen und erweiterbaren Datenstruktur abgebildet werden können.




**Abbildung 5.14:** Zusammenfassung der Methodik. Die integrative Analyse metabolischer Modelle beinhaltet a) Formalisierung ( $G_{Abstrakt}$  mit  $M$  Metabolit- und  $R$  Reaktionsknoten, verschiedenen Kantentypen:  $vi$  Verbrauch irreversibel,  $vr$  Verbrauch reversibel,  $pi$  Produktion irreversibel,  $pr$  Produktion reversibel und  $i$  Inhibition), b) Visualisierung in SBGN-PD, c) Ableitungen in verschiedene *spezifische Graphen*:  $G_{Kinetik}$  Dunkelgrün,  $G_{Petri-Netz}$  Hellgrün,  $G_{Stöchiometrie}$  Hellblau und zwei topologischen Graphen  $G_{Metabolit}$  und  $G_{Reaktion}$  Dunkelblau und d) Integration verschiedener Analyseergebnisse (Farben repräsentieren Analyseergebnisse, ermittelt unter Verwendung unterschiedlicher Modellierungsformalismen und deren *spezifische Graphen*).

Diese Datenstruktur ermöglicht die Ableitung spezieller Datenstrukturen für Analysen struktureller und dynamischer Eigenschaften in verschiedenen Modellierungsformalismen. Unterschiedliche Analyseergebnisse werden in die einheitliche Datenstruktur integriert und unter Verwendung einer einheitlichen Visualisierung mittels Interaktionstechniken explorierbar.

Das Potential der unterschiedlichen Modellierungsformalismen wird in der entwickelten Methodik, aufgrund der Integration, Interaktionstechniken und Visualisierung im Kontext des Modells, in einen Arbeitsablauf nutzbar. Die integrative Analyse reduziert die Schwächen der unterschiedlichen und voneinander unabhängig betrachteten Formalismen (siehe Tabelle 5.2 im Vergleich zu Tabelle 3.1 in Kapitel 3).

Zur Charakterisierung der Verhaltens- und Funktionsweisen biologischer Systeme ist keine vollständig detaillierte Beschreibung der Reaktionen durch kinetische Ratengleichungen und Parameter notwendig.

**Tabelle 5.2:** Integration unterschiedlicher Modellierungsformalismen metabolischer Modelle reduziert Schwächen der voneinander unabhängig betrachteten Formalismen (siehe zum Vergleich Abbildung 3.1 Kapitel 3). Reduzierte Schwächen sind durch gelbe Plus/Minus bzw. grüne Plus gekennzeichnet, wobei Plus/Minus eine leichte Verbesserung und Plus die Umkehr zum Vorteil darstellt.



Details	Quantitative Modelle		Qualitative Modelle		Größe
	Kinetische Modellierung	Petri-Netz Modellierung	Stöchiometrische Modellierung	Topologische Modellierung	
	<ul style="list-style-type: none"> <li>+ Detaillierte quantitative Beschreibung</li> <li>+ Quantitative Vorhersagen und dynamisches Verhalten</li> </ul>	<ul style="list-style-type: none"> <li>+ Quantitative Beschreibung</li> <li>+ Erweiterbar (eSPN, ...)</li> <li>+ Quantitative Vorhersagen und dynamisches Verhalten</li> </ul>	<ul style="list-style-type: none"> <li>+ Fließgleichgewichtsannahme</li> <li>+ Quantitative Vorhersagen</li> <li>+ Große Modelle</li> </ul>	<ul style="list-style-type: none"> <li>+ Nur topologische Information</li> <li>+ Große Modelle</li> </ul>	
	<ul style="list-style-type: none"> <li>+ Kinetische Ratengleichungen und Parameter benötigt</li> <li>+ Kleine Modelle</li> </ul>	<ul style="list-style-type: none"> <li>+ Erweiterungen ermöglichen quantitative Vorhersagen</li> <li>+ Erweiterungen erschweren qualitative Analysen</li> </ul>	<ul style="list-style-type: none"> <li>+ Statische Beschreibung</li> <li>+ Keine dynamischen Eigenschaften</li> </ul>	<ul style="list-style-type: none"> <li>+ Keine dynamischen Eigenschaften</li> </ul>	

Eine Approximation durch die stochastische Petri-Netz Analyse oder Flussbilanzanalyse ergibt unter Verwendung weniger detaillierter Daten eine Annäherung an die präzise Lösung. Weiterhin können somit auch große Modelle zur quantitativen Vorhersage genutzt werden. Die integrierten Analyseergebnisse finden in weiterführenden Analysen Verwendung. Die Anwendung der integrierten *Dead-End* Analyse oder Zentralitätsanalysen ermöglichen erweiterte Petri-Netze qualitativ zu analysieren. Zudem können qualitative Modelle mit einer statischen Beschreibung durch die integrative Analyse dynamische Eigenschaften hervorbringen.

Die Integration der verschiedenen metabolischen Modellierungsformalismen ermöglicht vielfältige und sich ergänzende Analysen, die ein tieferes Verständnis vom Metabolismus des zugrunde liegenden biologischen Systems ergeben. Diese Methodik ist in Form einer integrierten Anwendung dem *System Biology Metabolic Modell Framework* (kurz *SBM<sup>2</sup>-Framework*) realisiert und wird im folgenden Kapitel 6 beschrieben.



## 6 Realisierung

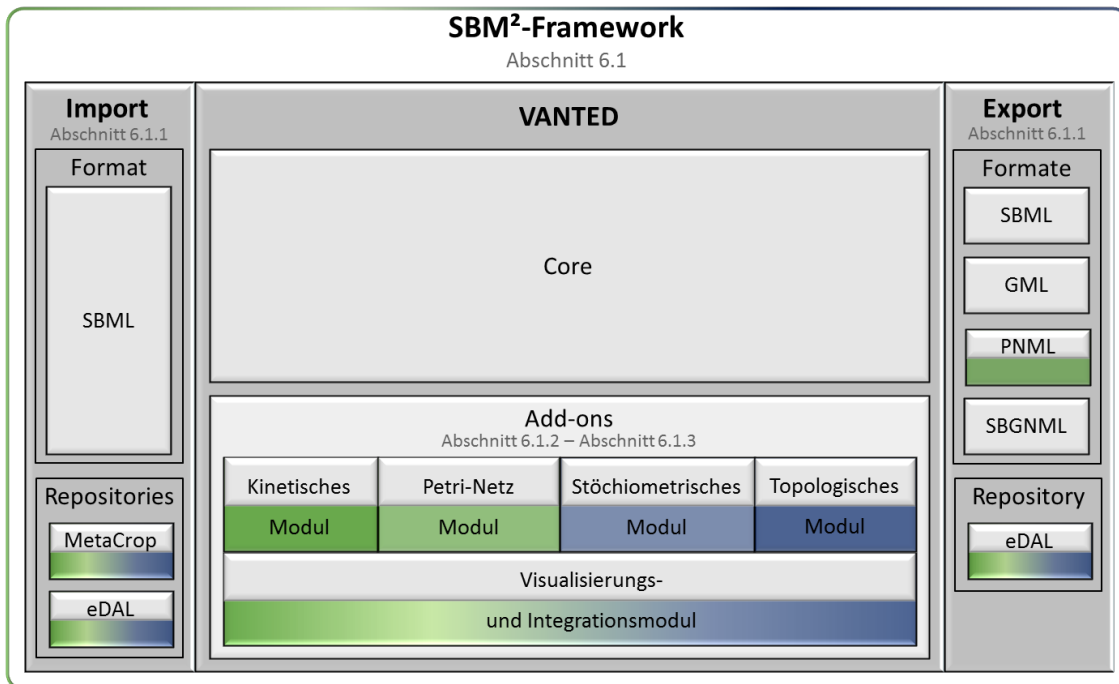
Zur metabolischen Modellierung existieren eine Vielzahl von Anwendungen, die bereits in Kapitel 4 untersucht wurden. Die Anforderungen an eine integrierte Anwendung, die (1) kinetische, Petri-Netz, stöchiometrische und topologische Modellierung, (2) einheitliche Visualisierung (*SBGN*), (3) Analyseergebnisse im Kontext des Modells und (4) Exploration durch Interaktionstechniken unterstützt, konnte in dieser Evaluierung nicht ermittelt werden. Die Entwicklung einer Methodik, die diese Anforderungen erfüllt, ist in Kapitel 5 erarbeitet worden. Auf dieser Grundlage wird die Realisierung in Form einer integrierten Anwendung, dem *System Biology Metabolic Model Framework* (kurz *SBM<sup>2</sup>-Framework*), im Folgenden beschrieben.

### 6.1 System Biology Metabolic Model Framework

Das entwickelte *SBM<sup>2</sup>-Framework* unterstützt den Anwender bei der metabolischen Modellierung. Es ermöglicht einen fließenden und interaktiven Ablauf der Modellierung und stellt eine Auswahl von vier verschiedenen Modellierungsformalismen bereit. Unterschiedliche metabolische Modelle werden einheitlich visualisiert und Analyseergebnisse in das Modell explorierbar integriert.

Das *SBM<sup>2</sup>-Framework* basiert auf dem Visualisierungs- und Analysewerkzeug *VANTED* (*Visualization and Analysis of Networks containing Experimental Data*) [128] (siehe Kapitel 4.1 Tabelle 4.5). *VANTED* ermöglicht unter anderem die Erstellung von Netzwerken, die Integration verschiedener biologischer Datentypen und die Exploration dieser Daten im Kontext von Netzwerken. Zudem finden verschiedene systembiologische Standards in der Visualisierung und dem Datenaustausch Anwendung. Als offenes System bietet *VANTED* eine Schnittstelle zur Erweiterung der Funktionalität. Diese Schnittstelle wird auch bei dem hier vorgestellten Framework verwendet, siehe Abbildung 6.1.

Das *SBM<sup>2</sup>-Framework* ist modular aufgebaut. Die Hauptkomponente ist *VANTED* mit dessen *Core*-Funktionalität und Import sowie Export Modulen. *VANTED-Core* stellt die Verarbeitung von Graphen bereit, die wie in der Methodik 5 beschrieben, die Basis des *SBM<sup>2</sup>-Framework* bilden. *VANTED*'s Import und Export Module ermöglichen die Verwendung verschiedener Austauschformate. Zur metabolischen Modellierung wird das Standard-Austauschformat metabolischer Modelle *SBML* importiert. Neben *SBML* wird der Export metabolischer Modelle in den Formaten *PNML*, *SBGNML* und *GML* angeboten (siehe Kapitel 2 Abschnitt 2.3.2). Weiterhin können Repositories zum Laden (*MetaCrop*, *eDAL*) und zur Ablage (*eDAL*) von Modellen verwendet werden. Die Hauptmodule sind Add-ons, die verschiedene Modellierungsformalismen implementieren (kinetische, Petri-Netz, stöchiometrische und topologische Modellierung, siehe Kapitel 3) und ein Add-on zur einheitlichen Visualisierung in *SBGN* (siehe Kapitel 2 Abschnitt 2.3.3.1). Diese Module werden im Folgenden vorgestellt.



**Abbildung 6.1:** Das *SBM<sup>2</sup>-Framework* importiert metabolische Modelle definiert in *SBML* und ermöglicht den Export in verschiedene Dateiformate. Modelle können aus Repositories (METACROP, eDAL) geladen und abgelegt (eDAL) werden. Basierend auf VANTED's Hauptimplementierung (VANTED-Core) besteht das *SBM<sup>2</sup>-Framework* aus verschiedenen Modulen zur Modellierung metabolischer Modelle, die in Form von Add-ons implementiert sind. Das *SBM<sup>2</sup>-Framework* verwendet zur einheitlichen Visualisierung eine Funktion des Add-ons *SBGN-ED* und ermöglicht die Integration der Analyseergebnisse in dieser Visualisierung. (Eigene Implementierungen, Weiterentwicklungen und inhaltliche Erweiterungen sind farblich gekennzeichnet.)

### 6.1.1 Import und Export metabolischer Modelle


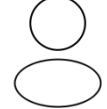
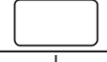

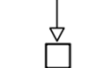

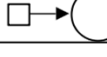
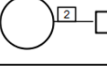
Das *SBM<sup>2</sup>-Framework* verwendet *SBML* [74] als Hauptaustauschformat. *SBML* erlaubt die Definition aller notwendigen Informationen für die vier bereitgestellten Modellierungsformalisten. Qualitative und quantitative metabolische Modelle, die in *SBML* definiert sind, werden importiert [96] und mittels des *SBM<sup>2</sup>-Framework* in die Datenstruktur des abstrakten Graphen (siehe Kapitel 5 Abschnitt 5.1.2) überführt. Diese Überführung erfolgt anhand der vorgegebenen Struktur entsprechend der *SBML* Spezifikation [136].

Aufgrund der zentralen Datenstruktur des *abstrakten Graphen* können die verwendeten metabolischen Modelle in den Formaten *SBML*, *PNML*, *GML* und *SBGNML* gespeichert werden. Der Export metabolischer Modelle in das Petri-Netz Format *PNML* [121] erfolgt über eine Transformation der *SBML* in die Petri-Netz Struktur entsprechend der *PNML* Spezifikation [69], siehe Tabelle 6.1.

Zum Abspeichern der Struktur, des Layouts und der gesamten Metainformationen (im modellintegrierte Analyseergebnisse) metabolischer Modelle wird das *GML* [70] Format bereitgestellt [128]. Der Export in *SBGNML* [158] erhält die Struktur und das Layout des Modells [32]. Des Weiteren ist VANTED, und somit das *SBM<sup>2</sup>-Framework* direkt mit dem METACROP Repository [140] verbunden. METACROP enthält detaillierte, manuell kurierte Informationen zum Stoffwechsel von ökonomisch bedeutenden Kulturpflanzen. Stoffwechselwege und metabolische Modelle können in den Formaten *SBML*, *SBGNML* und



**Tabelle 6.1:** Transformation von biologischem Wissen in *SBGN-PD* (Symbole und deren Bezeichner), *SBML* und *PNML* ist dargestellt.

Biologie	SBGN-PD Symbole	SBML	PNML
Kompartiment	 compartment	compartment	(sub-) model
Metabolit	 simple chemical / unspecified entity	species	place
Enzym	 macromolecule	modifier	place
Reaktion	 catalysis of a process	reaction with modifier	transition
Transport	 stimulation of a process	reaction with modifier	transition
Verbrauch	 consumption	reactant of reaction	input arc
Produktion	 production	product of reaction	output arc
Stöchiometrie	 cardinality of consumption / production arc	stoichiometry of reactants or products	arc weight
Quantitäten	kein Symbol	initial amount / initial concentration	discrete / continuous token
Kinetik	kein Symbol	kinetic law of reaction	transition rate

*GML* exportiert werden. Ein Weiteres mit *VANTED* verbundenes Repository für metabolische Modelle ist die *electronical Data Archive Library* (eDAL) [5]. Die Implementierung dieser Bibliothek als *VANTED* Add-on ermöglicht metabolische Modelle um zusätzliche Metadaten zu ergänzen und versionskontrolliert abzulegen. Der iterative Prozess der Modellierung (siehe Kapitel 2 Tabelle 2.6) wird somit durch das *SBM<sup>2</sup>-Framework* unterstützt.

### 6.1.2 Visualisierung metabolischer Modelle

Zur einheitlichen Visualisierung metabolischer Modelle wird eine Transformationsfunktion aus dem *VANTED* Add-on *SBGN-ED* [32] verwendet. *SBGN-ED* erlaubt die Erstellung und Bearbeitung der drei *SBGN* Teilsprachen *SBGN-PD*, *SBGN-ER* und *SBGN-AF* (siehe Kapitel 2 Abschnitt 2.3.3.1) und deren Validierung entsprechend der *SBGN* Spezifikationen. Ausgehend vom abstrakten Graphen (*SBML* Attribute) im *SBM<sup>2</sup>-Framework* werden metabolische Modelle durch die Transformationsfunktion *SBML2SBGN* in *SBGN-PD* einheitlich visualisiert, siehe Tabelle 6.1. Dazu werden *SBML* Attribute und (falls in *SBML* definiert) *SBO* Annotationen zur Transformation verwendet. Entsprechende Zuordnungen von *SBO* Annotationen zu *SBGN-PD* Symbolen, können der *SBGN-PD* Spezifikation entnommen werden [112].

Zusätzliche Eigenschaften metabolischer Modelle wie Quantitäten (z. B. Metabolitkonzentrationen, Reaktionsraten) werden nicht in *SBGN-PD* repräsentiert. In den Modulen zur Analyse werden diese Quantitäten definiert oder sind bereits in der Modelldefinition des Dateiformats hinterlegt, sodass der *abstrakte Graph* bereits damit angereichert ist.

### 6.1.3 Analyse metabolischer Modelle

Eine Auswahl aus vier verschiedenen Modellierungsformalisten unterstützt den Nutzer, metabolische Modelle vielfältig zu analysieren (siehe Kapitel 3). Ein importiertes Modell wird dazu ausgehend vom *abstrakten Graphen* und abhängig vom gewählten Formalismus in die *speziellen Graphen* zur Analyse abgeleitet (siehe Kapitel 5 Abschnitt 5.2).

#### 6.1.3.1 Kinetische Analyse

Die kinetische Analyse setzt detailliertes Wissen über Interaktionsmechanismen metabolischer Prozesse eines biologischen Systems voraus (siehe Kapitel 3 Abschnitt 3.2). Metabolitkonzentrationen und Reaktionskinetiken mit zugehörigen Parametern gehen in die Analyse ein. Kinetische Modelle werden am häufigsten für Zeitreihensimulationen und das Lösen des Fließgleichgewichtes verwendet [130].

Das kinetische Modul nutzt das VANTED Add-on *VisKi* [65], welches *SBML* definierte kinetische Modelle simuliert, auf Fließgleichgewicht prüft und visualisiert. *VisKi* stellt acht verschiedene numerische Lösungsverfahren aus der *Simulation Core Library* [83] für Zeitreihensimulationen bereit. Abhängig von den Eigenschaften des Modells kann zwischen den expliziten Einschrittverfahren (*Euler*, *Higham-Hall54*, *Runge-Kutta4*, *Dormand-Price54* und *Dormand-Price853*), dem impliziten Einschrittverfahren (*Rosenbrock*), den expliziten Mehrschrittverfahren (*Adams-Bashforth* und *Gragg-Bulirsch-Stoer*) und dem impliziten Mehrschrittverfahren (*Adams-Moulton*) gewählt werden.

Die Simulation wird mit den Parametern Startzeit, Endzeit und Schrittweite initialisiert. Die Simulationsergebnisse können zum einen in einem Zeitreihendiagramm zusammengefasst werden und zum anderen als separate Diagramme in das Modell integriert betrachtet werden. Das Zeitreihendiagramm enthält Kurven für alle Komponenten oder eine Auswahl der im Modell enthaltenen Komponenten wie die Metabolitkonzentrationen, Reaktionsraten und Parameterwerte über die Zeit. Die Auswahl von Modellkomponenten zur Diagrammdarstellung erfolgt anhand der Knoten im Modell oder über Checkboxen, wobei die Farben der Komponenten denen im Diagramm entsprechen. In der modellintegrierten Visualisierung enthält jeder Knoten ein entsprechendes Diagramm, z. B. erhält ein Metabolitknoten ein Zeitreihendiagramm mit seiner Konzentration über die Zeit. Diese Visualisierung unterstützt den Nutzer bei der Exploration, Interpretation und visuellen Analyse des Modells auf Systemebene.

Des Weiteren ermöglicht *VisKi* den Vergleich einer beliebigen Anzahl Simulationen mit z. B. geänderten Modellparametern modellintegriert. Dazu können die initialen Metabolitkonzentrationen und kinetische Parameter angepasst und eine erneute Simulation durchgeführt werden. Die berechneten Zeitreihendiagramme werden dann neben den zuvor integrierten Diagrammen hinzugefügt. Zusätzlich werden die Simulationsergebnisse in einer Tabelle zusammengetragen, die für weitere Untersuchungen exportiert werden kann.

*VisKi* bietet auch die Möglichkeit der Analyse des Fließgleichgewichtes. Stellt sich während einer Simulation für die im Modell enthaltenen Komponenten ein Fließgleichgewicht ein, wird es in einer Tabelle dokumentiert.

#### 6.1.3.2 Petri-Netz Analyse

Petri-Netze stellen einen mathematischen Formalismus zur Modellierung, Simulation und zur Analyse biologischer Systeme bereit. Abhängig von den integrierten quantitativen

Daten können Petri-Netze qualitativ und quantitativ analysiert werden (siehe Kapitel 3 Abschnitt 3.3).

Das Petri-Netz Modul nutzt eine erweiterte Version des VANTED Add-ons PETRINET [121, 67] zur Validierung, Simulation und Analyse erweiterter *P/T-Netze* (ePNs) und erweiterter *SPNs* (eSPNs). Die Erweiterung bezieht sich auf die Möglichkeit, die folgenden Parameter und Kantentypen zu nutzen. Es können kontinuierliche Marken (zum Modellieren von Metabolitkonzentrationen), kontinuierliche Kantengewichte (zum Modellieren von Stöchiometrien aus reellen Zahlen), kontinuierliche Kapazitäten (zum Modellieren von limitierten Metabolitkonzentrationen) und Inhibitoranten (zur Modellierung von Inhibitionen) und für eSPNs gewichtete stochastische Schaltraten (zur Modellierung von Reaktionsraten) verwendet werden (Details siehe Kapitel 3 Abschnitt 3.3).

Das Petri-Netz Modul ermöglicht qualitative und quantitative Analysen metabolischer Modelle. Die qualitativen Analysen beinhalten die Berechnung von Invarianten und der Erreichbarkeit, um spezifische Modelleigenschaften zu prüfen. Die Interaktionstechnik *Brushing & Linking* (siehe Kapitel 2 Abschnitt 2.3.3.2) erlaubt die Betrachtung der berechneten Invarianten, integriert im metabolischen Modell. Die quantitative Analyse zur Untersuchung des dynamischen Systemverhaltens eines metabolischen Modells erfolgt über die Petri-Netz Simulation. Hier kann zwischen zwei Modi unterschieden werden: der nicht-deterministischen Simulation (ePN) und der stochastischen Simulation (eSPN). Das Verhalten biologischer Systeme wird von stochastischen Gesetzmäßigkeiten beeinflusst, sodass sich die stochastische Simulation für metabolische Modelle eignet [55].

Die Simulationsergebnisse werden modellintegriert visualisiert. Während der Simulation wird mittels der Interaktionstechnik *Animation* der Verlauf der Änderung von Metabolitkonzentrationen schrittabhängig dargestellt. Die Anzahl der zu jedem Schritt aktuellen Metabolitkonzentration wird durch die Anzahl Marken im Modell visualisiert. Die Aktivität einer Reaktion wird durch die Markierung dieser zu einem Schritt angegeben. Am Ende der Simulation erhält jeder Knoten ein entsprechendes Diagramm, z. B. erhält ein Metabolitknoten ein Diagramm mit der Konzentration über die Simulationsschritte. Diese Visualisierung unterstützt den Nutzer bei der Exploration, Interpretation und visuellen Analyse des Modells auf Systemebene.

Des Weiteren ermöglicht PETRINET den Vergleich einer beliebigen Anzahl Simulationen mit z. B. geänderten Modellparametern. Dazu können die initialen Metabolitkonzentrationen und Reaktionsraten angepasst werden. Die berechneten Diagramme werden dann neben den zuvor integrierten Diagrammen hinzugefügt. Zusätzlich werden die Simulationsergebnisse in einer Tabelle zusammengetragen, die für weitere Untersuchungen verwendet werden kann.

### 6.1.3.3 Stöchiometrische Analyse

Im Allgemeinen ist detailliertes Wissen zu biochemischen Reaktionen nicht präzise, einheitlich und vollständig [8]. In diesem Fall ist die stöchiometrische Analyse metabolischer Modelle geeignet, die aufgrund der Annahme des Fließgleichgewichtes, Aussagen über das biologische System treffen kann.

Das stöchiometrische Modul nutzt eine erweiterte Version des VANTED Add-ons FBASIMVIS [62] zur strukturellen und optimierungsbasierten Analyse metabolischer Modelle. Zum einen wird das metabolische Modell strukturell mit der *Dead-End* Analyse und zum anderen mit den optimierungsbasierten Analysen *Gap-Finding*, Flussbilanzanalyse (*FBA*), Flussvariabilitätsanalyse (*FVA*), Robustheitsanalyse (*RA*) und Knockout-Analyse

(*KA*) untersucht (Details siehe Kapitel 3 Abschnitt 3.4). Zur Verifikation metabolischer Modelle dienen die *Dead-End* und *Gap-Finding* Analysen. Die optimierungsbasierten Analysen (*FBA*, *FVA*, *RA* und *KA*) treffen aufgrund zusätzlicher Bedingungen und der Angabe einer Zielfunktion, Aussagen über das metabolische Potential eines Modells.

Die Analyseergebnisse werden in das metabolische Modell integriert. Die Ergebnisse der Verifikation sind *Dead-End* Metabolite, welche im Modell mittels einer Färbung angezeigt werden. Flussverteilungen sind Ergebnisse der *FBA*, die zu einem optimalen metabolischen Ertrag führen. Diese Flussverteilungen werden durch eine Skalierung der Kantenbreiten im Modell dargestellt. Je breiter die Kanten, desto größer ist der Metabolitfluss durch die entsprechende Reaktion. *FBASIMVIS* ermöglicht den Vergleich von Flussverteilungen mit geänderten Modellparametern. Unterstützt durch die Interaktionstechnik *Animation* können die Parameter schrittweise variiert werden, sodass die davon abhängigen Flussverteilungen (entsprechend der gewählten Schrittweite) dynamisch angepasst werden. Die Sensitivitätskurven werden als Resultate der Robustheitsanalyse in Form eines Kurvendiagrammes, die Sensitivitätswerte (Resultat der Knockout-Analyse) und die minimalen und maximalen Flusswerte (Resultat Flussvariabilitätsanalyse) als Balkendiagramme in den Reaktionsknoten visualisiert.

#### 6.1.3.4 Topologische Analyse

Die topologische Analyse untersucht die Struktur und zentrale Elemente des metabolischen Modells (siehe Kapitel 3 Abschnitt 3.5). Der Nutzer kann dazu das Modell basierend auf den *Metabolit-* oder *Reaktionsgraphen* analysieren. Die Analyse des kürzesten Weges zwischen zwei Metaboliten oder Reaktionen wird mit Hilfe der Funktionalität in *VANTED-Core* berechnet.

Das topologische Modul nutzt weiterhin das *VANTED* Add-on *CENTILIB* [64] zur Zentralitätsanalyse, um zentrale Elemente eines metabolischen Modells zu klassifizieren. In Bezug auf metabolische Modelle findet die Zentralitätsanalyse die wichtigsten a) Metabolite in einem *Metabolitgraphen* oder b) Reaktionen in einem *Reaktionsgraphen*, die essentiell für das Verständnis biologischer Prozesse sind. *CENTILIB* stellt 17 verschiedene Knotenzentralitäts- und vier Graphzentralitätsanalysen bereit, die geordnete Listen von Zentralitätswerten ermitteln. Die Interaktionstechnik *Brushing & Linking* (siehe Kapitel 2 Abschnitt 2.3.3.2) erlaubt die berechneten Zentralitätswerte integriert im metabolischen Modell zu untersuchen.

## 6.2 Zusammenfassung

Das entwickelte *SBM<sup>2</sup>-Framework* realisiert die erarbeitete Methodik aus Kapitel 5. Das Potential der unterschiedlichen Modellierungsformalismen wird mit dem *SBM<sup>2</sup>-Framework* aufgrund der Integration, Interaktionstechniken und Visualisierung im Kontext des Modells in einem durchgängigen Arbeitsablauf nutzbar. Die Anforderungen an eine integrierte Anwendung, die (1) kinetische, Petri-Netz, stöchiometrische und topologische Modellierung, (2) einheitliche Visualisierung (*SBGN*), (3) Analyseergebnisse im Kontext des Modells und (4) Exploration durch Interaktionstechniken unterstützt, sind somit erfüllt. Die Anwendung des *SBM<sup>2</sup>-Framework* wird im folgenden Kapitel 7 anhand verschiedener Vorgehensweisen präsentiert. (Das *SBM<sup>2</sup>-Framework*, inklusive eines Tutorials, ist auf der beiliegenden CD enthalten.)

## 7 Anwendung

Die entwickelte Methodik zur integrativen Analyse metabolischer Modelle ist als *System Biology Metabolic Model Framework* (kurz *SBM<sup>2</sup>-Framework*) realisiert. In diesem Kapitel wird anhand zwei verschiedener Vorgehensweisen die Ableitung struktureller und dynamischer Eigenschaften metabolischer Modelle erarbeitet. Zunächst werden verschiedene, bereits publizierte, metabolische Modelle zur Modellierung herangezogen, um spezifische Analysen durchzuführen. Dieser Abschnitt dient dem Nachweis der Verwendbarkeit der entwickelten Methodik zur Analyse spezifischer Modelle in einem Computerprogramm. Anschließend wird die integrative Analyse eines publizierten Kartoffel-Modells und zweier selbst entwickelter Modelle der Maispflanze durchgeführt. Das Vorgehen in diesem Kapitel ist in Abbildung 7.1 dargestellt.

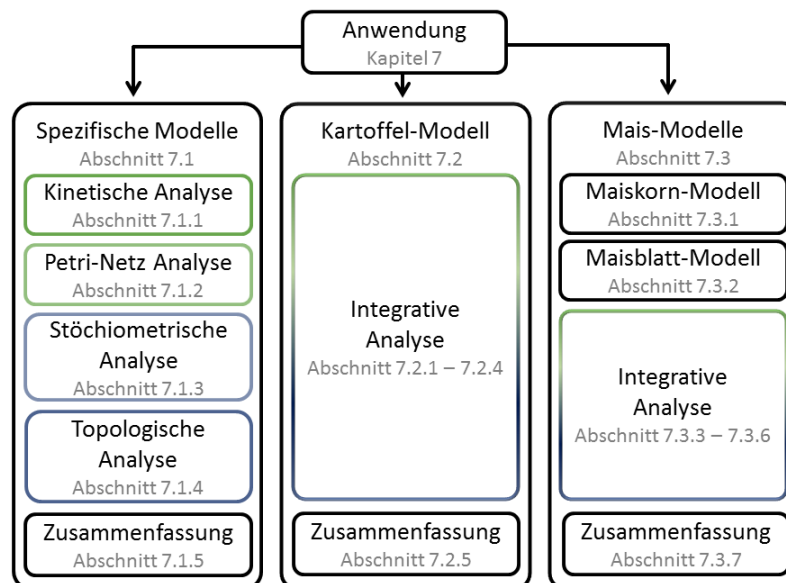


Abbildung 7.1: Struktur des Kapitels Anwendung.

### 7.1 Anwendung verschiedener spezifischer Modelle

Abhängig von der zu lösenden Fragestellung, dem biochemischen Wissen und der Verfügbarkeit experimenteller Daten zu Interaktionsmechanismen, können biologische Systeme durch verschiedene metabolische Modelle repräsentiert werden. Folglich entstehen große, meist qualitative Modelle und kleine, eher quantitative Modelle [153]. Unterschiedliche Modellgrößen und enthaltene Details ermöglichen die Analyse struktureller und dynamischer Eigenschaften unter Verwendung verschiedener Modellierungsfomalismen. Quantitative und qualitative Modelle können mit der implementierten Methodik mittels unterschiedlicher Modellierungsfomalismen analysiert und exploriert werden.

### 7.1.1 Kinetische Analyse

Das von Rohwer und Botha [131] rekonstruierte kinetische Modell der Saccharose-Anreicherung im Zuckerrohrhalm beinhaltet eine detaillierte Beschreibung zu Reaktionsmechanismen. Es wird zur kinetischen Analyse verwendet. Das Modell repräsentiert 13 Metabolite und 11 Reaktionen. Davon beschreiben drei Reaktionen Transportprozesse zwischen der Umgebung, dem Zytosol und der Vakuole. Zur Untersuchung des zeitlichen Ablaufs der biochemischen Reaktionen wird die Simulation mit der Euler Methode durchgeführt (Zeitspanne 0 - 4.000, Schrittweite 1). Die entstehenden Metabolitkonzentrationen und Reaktionsraten über die Zeit werden in Form von Zeitreihendiagrammen in das Modell integriert, siehe Abbildung 7.2 (Zeitreihendiagramme links in den Knoten).

Die Besonderheit der kinetischen Analyse ist die Vergleichsmöglichkeit von Analyseergebnissen mit veränderten Parametern. Der Einfluss einer erhöhten Fruktose-Verfügbarkeit ( $Frux = 50$  mM) auf das dynamische Verhalten des Modells wird in einer zweiten Simulation ermittelt, siehe Abbildung 7.2 (Zeitreihendiagramm rechts in den Knoten). Die Diagramme auf der linken Seite basieren auf der gemessenen und die der rechten Seite auf der erhöhten Fruktose-Verfügbarkeit. Die Kurvenverläufe zeigen das Konvergieren bis zum Fließgleichgewicht in der integrierten Visualisierung. Ein vergrößerter Ausschnitt des Metaboliten Saccharose (*Suc*) ist in Abbildung 7.2 b) dargestellt. Bei einer erhöhten Fruktose-Verfügbarkeit wird mehr Saccharose gebildet, die zum Speichern in die Vakuole transportiert wird.

### 7.1.2 Petri-Netz Analyse

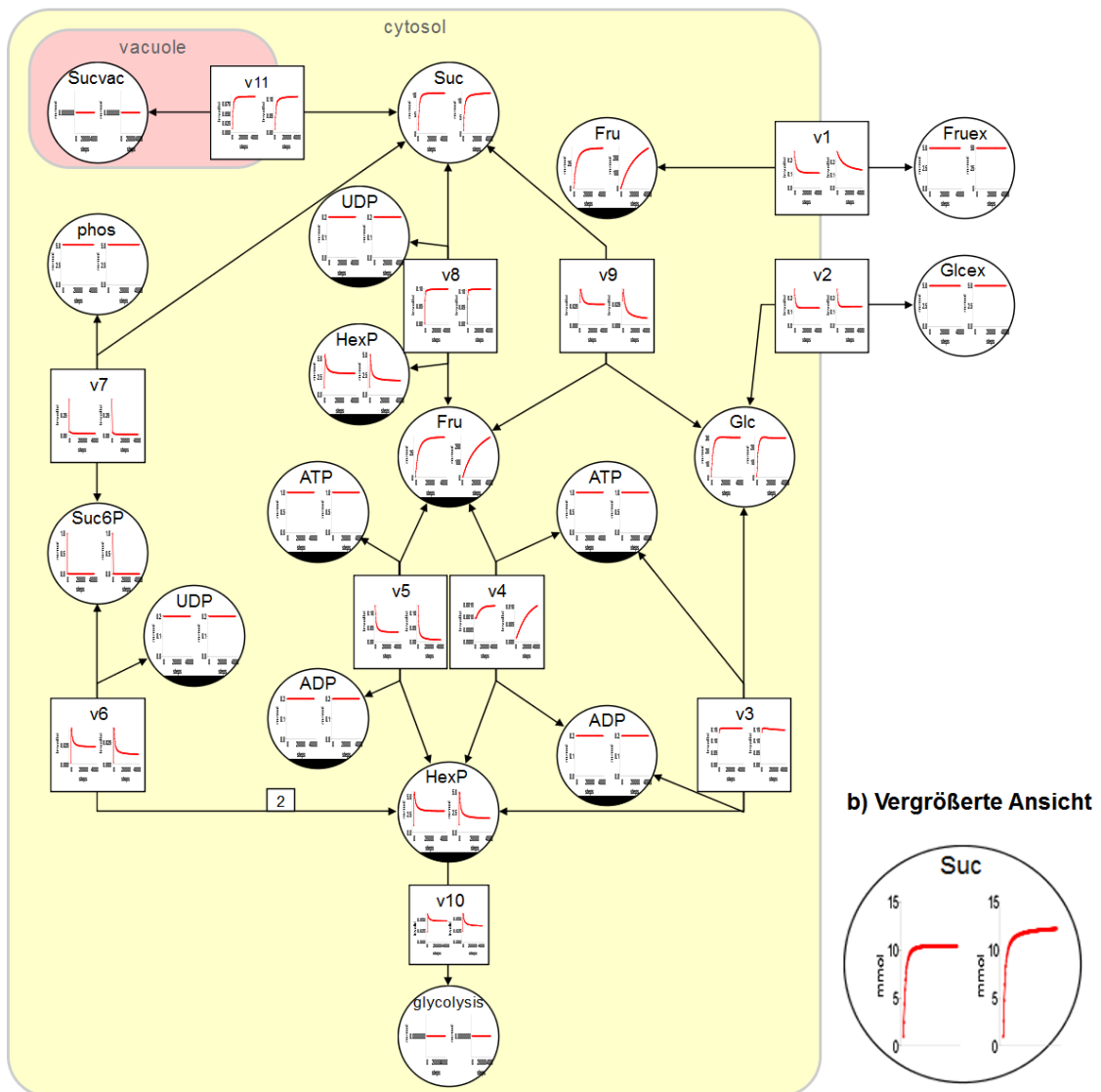
Petri-Netz Modelle können verschiedenen Typs sein. Sie werden qualitativ als auch quantitativ analysiert. Der Saccharose-Stoffwechsel im Speichergewebe einer einkeimblättrigen Pflanze dient der Invariantenanalyse, wie durch Schuster und Junker [143] beschrieben. Das Modell besteht aus 21 Metaboliten und 21 Reaktionen. Davon sind sieben Transportprozesse zwischen Umgebung, Zytosol und Plastid.

Die Berechnung der P- und T-Invarianten wird zur Prüfung spezifischer Modelleigenschaften und somit zur Validierung durchgeführt. P-Invarianten werden als Erhaltungsregeln eines metabolischen Modells interpretiert. Die T-Invarianten beschreiben Reaktionen, die Hauptstoffwechselwege bilden. Die Beziehung zwischen der Struktur des Petri-Netzes und dem dynamischen Verhalten wird dargestellt. Die Invariantenanalyse des Saccharose-Stoffwechsel-Modells resultiert in vier P-Invarianten und fünf T-Invarianten. Eine der P-Invarianten ist in Abbildung 7.3 durch Angabe der Gewichte (Zahlen in Magenta) im Modell integriert dargestellt. Diese P-Invariante zeigt die Stoffhaltung der Phosphatgruppen, die über verschiedene Metabolite gebunden vorliegen.

### 7.1.3 Stöchiometrische Analyse

Die stöchiometrische Analyse untersucht metabolische Modelle unter Optimierung einer Zielfunktion im Fließgleichgewicht, zum einen auf Validität und zum anderen auf die metabolische Flussverteilung. Mit dem Ziel, die optimale Flussverteilung unter Maximierung der Wachstumsrate zu untersuchen, wird die Flussbilanzanalyse am Beispiel des von Grafahrend-Belau und Koautoren [63] entwickelten Gerstensamen-Modells durchgeführt. Das Modell besteht aus 238 Metaboliten und 257 Reaktionen.

## a) Kinetische Analyse des Zuckerrohrhalm-Modells



**Abbildung 7.2:** Kinetische Analyse der Saccharose-Anreicherung im Zuckerrohrhalm-Modell. a) Jeweils zwei Diagramme in den Knoten visualisieren Metabolitkonzentrationen und Reaktionsraten über die Zeit (links mit initialer und rechts mit erhöhter Fruktose-Verfügbarkeit). b) Die vergrößerte Ansicht der Saccharose (*Suc*). (Eine hochauflösende Version der Abbildung ist auf der beiliegenden CD enthalten.)

Davon sind 65 Transportprozesse, die den Transport zwischen der Umgebung, dem Zytosol, dem Mitochondrium und dem Plastiden ermöglichen.

Unter Angabe der Saccharose-Aufnahmerate von  $8 \mu\text{M/gDW/h}$  und einer Sauerstoff-Aufnahmerate von  $4 \mu\text{M/gDW/h}$  (Sauerstoffmangel-Bedingung bei der Entwicklung des Gerstensamens) wird die optimale Flussverteilung berechnet. In Abbildung 7.4 fasst eine Überblickskarte die optimale Flussverteilung unter der Sauerstoffmangel-Bedingung zusammen. Die Stärkebiosynthese und der Citratzyklus erzeugen die wichtigen Biomassenbestandteile und sorgen für die in der Literatur beschriebene Wachstumsrate.





Umwandlung von Nitrit in Ammonium und wird aufgrund der Verbindung der beteiligten Metabolite mit vielen Reaktionen verknüpft.

### 7.1.5 Zusammenfassung

Die Anwendung verschiedener spezifischer Modelle mit unterschiedlichen qualitativen und quantitativen Eigenschaften weist die Verwendbarkeit der Methodik zur Analyse in speziellen Modellierungsformalismen nach. Anstelle mehrere Computerprogramme zu verwenden, dient das *SBM<sup>2</sup>-Framework* der Analyse und integrierten Visualisierung verschiedener Modelle.

## 7.2 Integrative Anwendung eines metabolischen Modells der Kartoffel

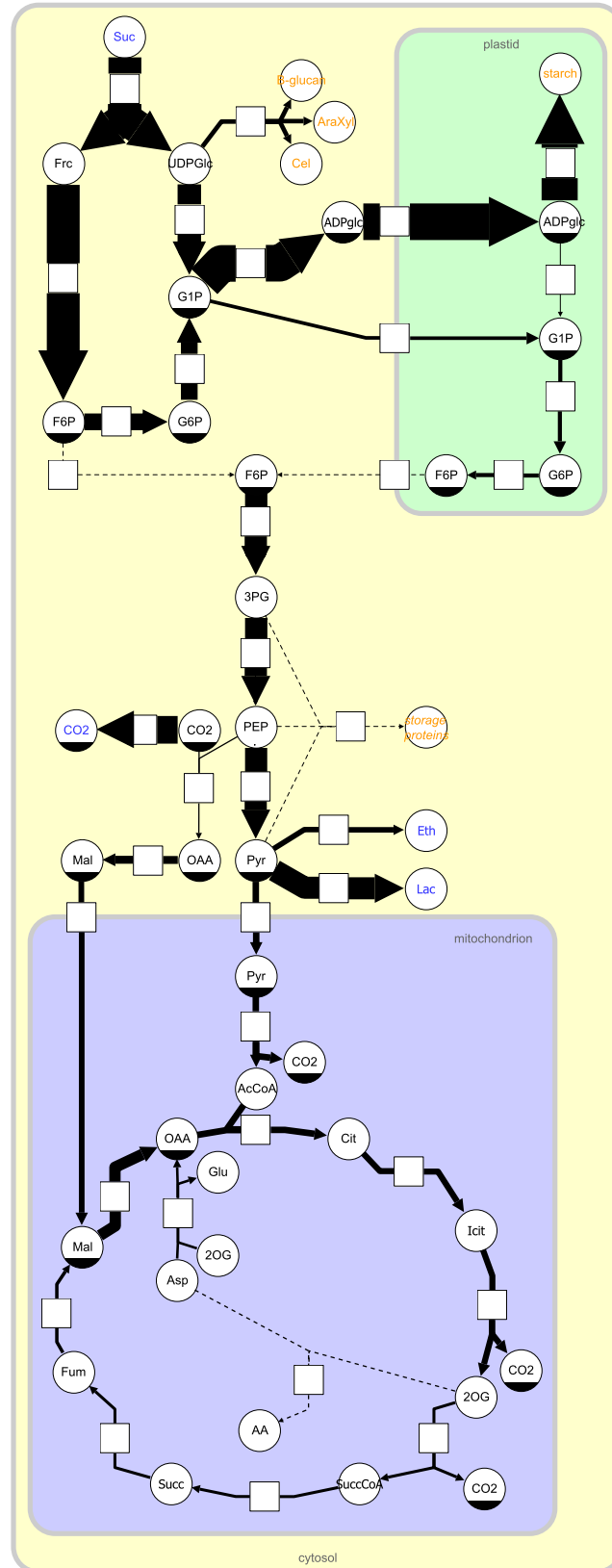
Die Kartoffel hat aufgrund ihres Hauptbestandteiles (Stärke) eine große Bedeutung als Nahrungsmittel und zur Herstellung von z. B. Kraftstoff und Textilien. Deshalb ist ein Hauptziel der Pflanzenzüchtung, die Verteilung von Biomasse innerhalb der Pflanze zugunsten der erntebaren Pflanzenteile zu verbessern. Das homogene Gewebe der Kartoffel ermöglicht die Untersuchung des Saccharoseabbaus. Dieser Stoffwechselweg beschreibt den Hauptmetabolitfluss in der Kartoffel von der Saccharose zur Stärke [53]. Nahezu alle Gene dieses Stoffwechselweges sind bereits bekannt und bilden somit die Voraussetzung der Rekonstruktion des metabolischen Modells der Kartoffel.

Zur Ausschöpfung aller in dieser Arbeit integrierten Modellierungsformalismen wird ein kinetisches Modell verwendet. Es beinhaltet eine detaillierte Beschreibung der Reaktionsmechanismen mit Parametern und initialen Metabolitkonzentrationen. Das von Junker [77] entwickelte kinetische Modell des Saccharoseabbaus in der sich entwickelnden Kartoffelknolle (kurz Kartoffel-Modell) bildet die Basis der folgenden Analysen. Es besteht aus 15 Reaktionen und 17 Metaboliten, die sich im Zytosol befinden, siehe Abbildung 7.5.

Der Stoffwechselweg beginnt mit dem Abbau der Saccharose (*Suc*) über verschiedene Reaktionen in Hexosephosphate (z. B. Glukose-6 Phosphat, *G6P*), die in die Glykolyse (*Glyc*) und Stärkebiosynthese (*StaSy*) eingehen. Zur Modellierung des Energieverbrauchs wird eine zusätzliche Reaktion (*ATPcons*) hinzugefügt. Die Reaktionen *Glyc*, *StaSy* und *ATPcons* werden zur Vereinfachung als zusammengefasste Reaktionen modelliert. Transportprozesse, die im Detail noch nicht verstanden sind und zusätzliche Kompartimente werden somit vermieden. Zur Darstellung der Umgebung werden die Transportreaktionen Saccharoseimport (*Imp*) und Stärkeexport (*Exp*) hinzugefügt. Diese Reaktionen werden in der kinetischen Analyse nicht betrachtet, finden jedoch in den folgenden Analysen Verwendung. Eine detaillierte Beschreibung der kinetischen Ratengleichungen und zugehörigen Parameter kann in der Dissertation von Junker [77] nachgelesen werden.

### 7.2.1 Kinetische Analyse

Das Kartoffel-Modell wird zur Untersuchung des zeitlichen Ablaufs der biochemischen Reaktionen mit der Euler Methode kinetisch analysiert (Zeitspanne 0 - 6.000, Schrittweite 0,1). Die entstehenden Metabolitkonzentrationen und Reaktionsraten über die Zeit werden in Form von Zeitreihendiagrammen in das Modell integriert, siehe Abbildung 7.5 (Zeitreihendiagramme links in den Knoten).



**Abbildung 7.4:** Die Flussbilanzanalyse (unter Sauerstoffmangel) des Gerstensamen-Modells resultiert in einer optimalen Flussverteilung. Die Überblickskarte visualisiert durch Kantenskalierungen, entsprechend der berechneten Flusswerte, das Ergebnis. Metabolite, die direkt in die Biomassebildung eingehen, sind Orange gekennzeichnet und Metabolite, die aufgenommen oder abgegeben werden, sind Blau gekennzeichnet. (Eine hochaufgelöste Version der Abbildung ist auf der beiliegenden CD enthalten.)

Das Kartoffel-Modell konvergiert nach einer Einschwingphase von etwa 4.600 zum Fließgleichgewicht mit Metabolitkonzentrationen, die in der Literatur [77] beschreiben sind.

Zur Untersuchung einer gesteigerten Stärkebiosynthese wird die Überexpression der Invertase (*Inv*) im Kartoffel-Modell nach Junker [77] angepasst. Die Überexpression wird durch die Erhöhung des Parameters  $v_{max}$  auf  $0,06 \text{ mM}/\text{FW}/\text{s}$  erreicht. Die Ergebnisse der kinetischen Analyse sind im Vergleich mit den vorhergehenden Resultaten in Abbildung 7.5 (Zeitreihendiagramme rechts in den Knoten) dargestellt.

Der Effekt der Überexpression auf die Metabolitkonzentrationen und Reaktionsraten über die Zeit wird durch eine länger andauernde Einschwingphase bis zum Fließgleichgewicht sichtbar. Die Diagramme sind der vorhergehenden Analyse ähnlich. Aber die Kurven verlaufen y-Achsen verschoben. Die Metabolite Glukose (*Glc*) und Fruktose (*Frc*) als auch die Reaktionen der Invertase (*Inv*) und Saccharose Synthase (*SuSy*), zeigen in der kinetischen Analyse der Überexpression Effekte, die in der Literatur [77] beschrieben sind. *Inv* ist anstelle von 3% nur für 85% des Saccharoseabbaus verantwortlich, sodass auch die Konzentrationen von *Glc* und *Frc* erhöht sind. Eine gesteigerte Bildung von Stärke (*Starch*) über die Stärkebiosynthese Reaktion (*StaSy*) konnte durch die Invertaseüberexpression erzielt werden, siehe Abbildung 7.5 b).

### 7.2.2 Petri-Netz Analyse

Die Petri-Netz Analyse des Kartoffel-Modells, im Hinblick auf Invarianten, wird detailliert von Koch und Koautoren in [89] zur Validierung des Modellverhaltens beschrieben. Als Beispiel wird einer der Hauptstoffwechselwege, den eine T-Invariante beschreibt, in Abbildung 7.9 durch Angabe der Gewichte (Schaltzähler in Magenta) dargestellt. Diese T-Invariante beschreibt das Spalten der Saccharose (*Suc*) durch die Invertase (*Inv*). Die entstehenden Hexosephosphate gehen in die Glykolyse und Stärkebiosynthese ein.

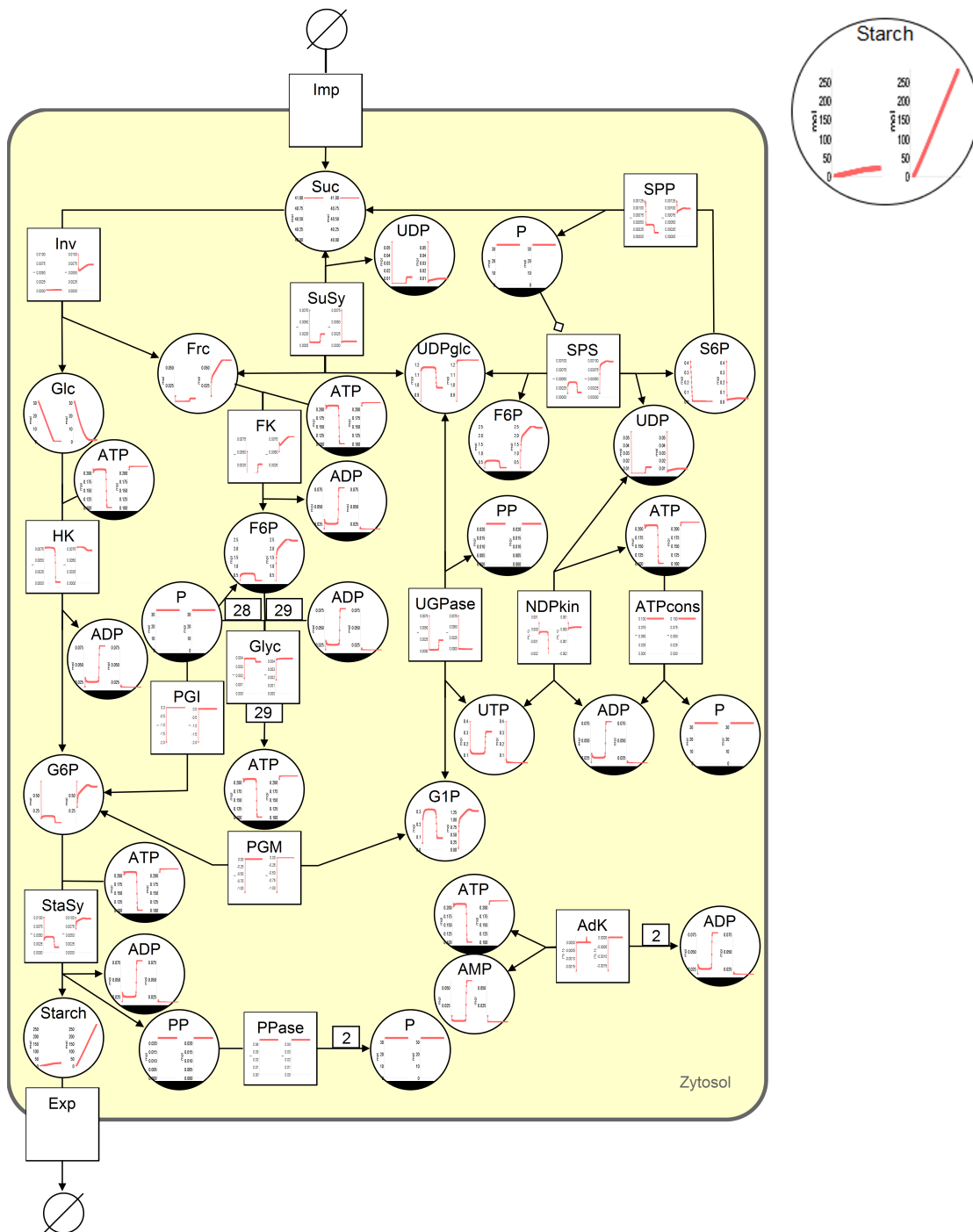
Ein anderer Ansatz, das dynamische Verhalten des Kartoffel-Modells zu analysieren, ist die stochastische Petri-Netz Simulation. Im Gegensatz zur Erreichbarkeits- und Invariantenanalyse [89], die die Ergebnisse für den Anwender aufbereiten, bedarf die Auswertung von Simulationsergebnissen der Interpretation des Anwenders. Zum Vergleich mit der kinetischen Analyse, die konvergierende Kurvenverläufe bis zum Fließgleichgewicht zeigt, soll der Einfluss der Stochastik auf die biochemischen Prozesse durch die stochastische Simulation betrachtet werden. Das Ziel dieser Simulationen ist das dynamische Verhalten des Kartoffel-Modells aus wenigen, detaillierten Daten (keine Reaktionsmechanismen und Regulation) abzuleiten.

Die Simulation benötigt die Angabe von Metabolitkonzentrationen für die Anfangsmarkierung des Petri-Netz Modells. Dazu werden die initialen Metabolitkonzentrationen des kinetischen Modells verwendet. Zur Beobachtung der Stärkebiosynthese während der Simulation wird die Exportreaktion (*Exp*) blockiert, siehe Abbildung 7.6 (rotes Kreuz im Knoten *Exp*). Zusätzlich werden alle Reaktionen mit den Reaktionsraten im Fließgleichgewicht des kinetischen Modells gewichtet. Damit erhalten Reaktionen mit einer höheren Reaktionsrate eine größere Wahrscheinlichkeit in einem Simulationsschritt ausgeführt zu werden (zu schalten) als solche mit geringeren Raten. In Abbildung 7.6 ist das Ergebnis der stochastischen Simulation nach 200.000 Schritten mit zehn Wiederholungen dargestellt.

Nach 200.000 Schritten produziert die Stärkebiosynthese Reaktion (*StaSy*) im Durchschnitt der zehn Wiederholungen  $9.073 \mu\text{M}$  Stärke. Die Reaktion *StaSy* beschreibt eine konstant aufsteigende Kurve.

a) Kinetische Analyse des Kartoffel-Modells

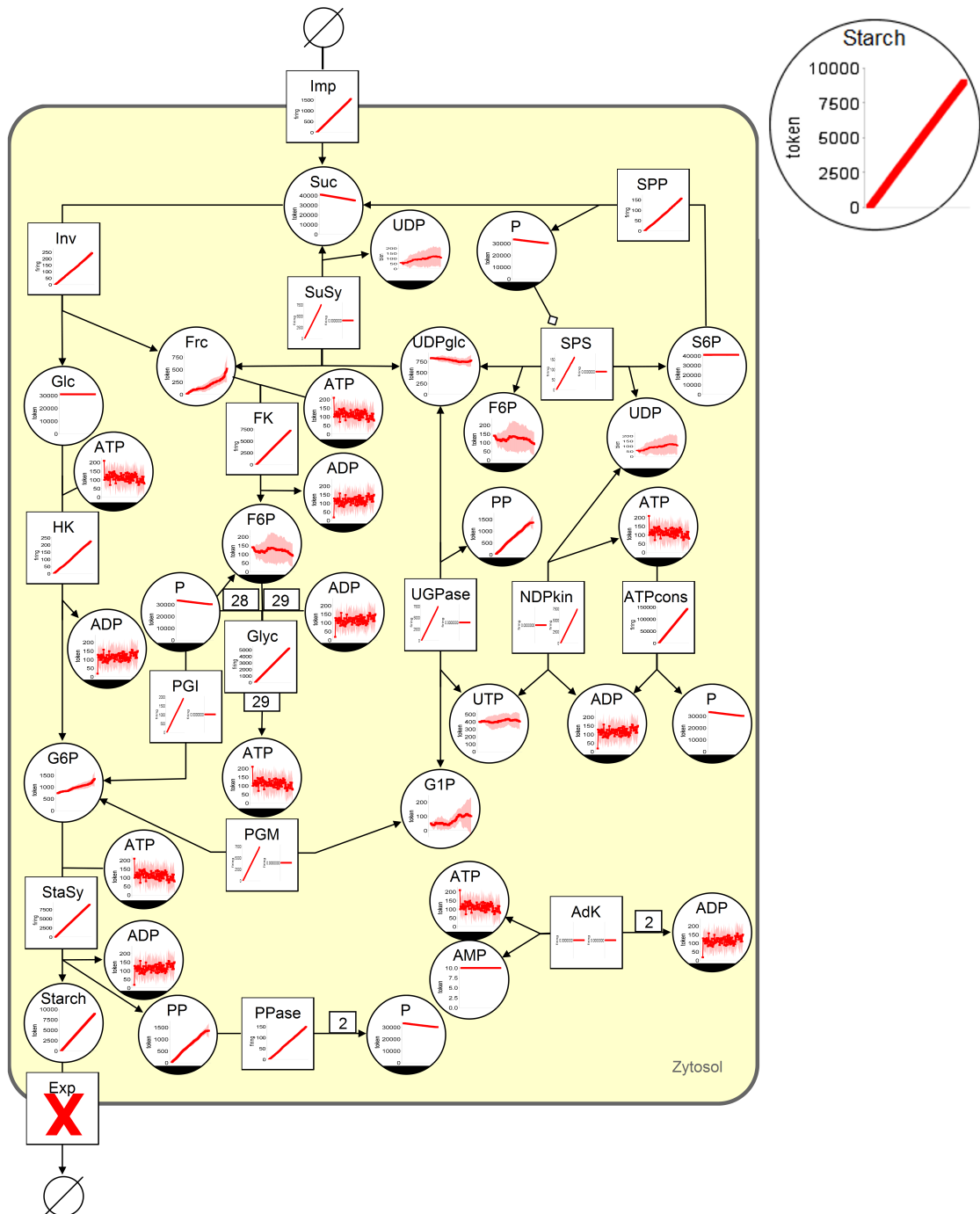
b) Vergrößerte Ansicht



**Abbildung 7.5:** Kinetische Analyse des Saccharoseabbaus in der sich entwickelnden Kartoffel. a) Jeweils zwei Diagramme in den Knoten visualisieren Metabolitkonzentrationen und Reaktionsraten über die Zeit (links mit initialer und rechts mit Überexpression der Invertase (*Inv*)). b) Die vergrößerte Ansicht der Stärke (*Starch*). (Eine hochauflösende Version der Abbildung ist auf der beiliegenden CD enthalten.)

## a) Stochastische Simulation des Kartoffel-Modells

## b) Vergrößerte Ansicht



**Abbildung 7.6:** Visualisierung der Ergebnisse zur stochastischen Petri-Netz Analyse des Saccharoseabbaus in der sich entwickelnden Kartoffel. a) Diagramme in den Knoten visualisieren Metabolitkonzentrationen und Reaktionsraten über 200.000 Schritte mit zehn Wiederholungen. Die Durchschnittswerte der Wiederholungen werden in der roten Kurve und die Abweichung vom Durchschnitt wird durch die hellroten Flächen ober- und unterhalb der Durchschnittskurve dargestellt. Reaktionen mit zwei Diagrammen zeigen die Raten der Hin- (links) und Rückreaktion (rechts) reversibler Reaktionen. Die Exportreaktion wird blockiert (rotes Kreuz im Knoten *Exp*), um die Biosynthese der Stärke zu betrachten. b) Die vergrößerte Ansicht der Stärke (*Starch*). (Eine hochauflösende Version der Abbildung ist auf der beiliegenden CD enthalten.)

Die stetige Aktivität dieser Reaktion wird somit gezeigt (Aufsummierung der Schaltrate über die Schritte). Nach 200.000 Schritten hat *StaSy* im Durchschnitt 8.866 mal geschaltet. Insgesamt laufen alle Reaktionen im Rahmen der angegebenen Rate ab, sodass die Kurvenverläufe der Reaktionen aufsteigend sind und sich die Werte am Ende der Simulation unterscheiden. Die Metabolite beschreiben verschiedene Kurvenverläufe. Neben *Starch* steigt auch die Konzentration des Pyrophosphates (*PP*) über die Schritte an, da *PP* auch in der *StaSy* gebildet wird. Die Konzentrationen unterscheiden sich am Ende der Simulation um den Faktor 6 ( $PP = 1.576 \mu M$ ), denn *PP* wird in zwei weiteren Reaktionen verbraucht. *Suc* und Phosphat (*P*) beschreiben einen absteigenden Kurvenverlauf. Beide Metabolite werden mit höheren Raten verbraucht als gebildet. Die energiereichen Verbindungen *ATP* und *ADP* nehmen an vielen Reaktionen teil und stellen oszillierende Kurvenverläufe dar. Der Grund liegt hauptsächlich an der Glykolyse Reaktion (*Glyc*), die beim schalten 29 *ADP* verbraucht und 29 *ATP* produziert. Die Kurvenverläufe der Hexosen (z. B. *Glc*) und Hexosephosphate (z. B. *G6P*) unterscheiden sich in den zehn Wiederholungen der Simulation. Dieses Verhalten wird durch die Abweichung der Konzentrationen vom Durchschnitt (hellrote Flächen ober- und unterhalb der Durchschnittskurve) sichtbar. Im Durchschnitt schwanken die Konzentrationen sehr gering, sodass über die Schritte eine Balance zwischen Produktion und Verbrauch entsteht. Fruktose (*Frc*) und *G6P* zeigen eher eine aufsteigende Tendenz.

Die stochastische Simulation des Kartoffel-Modells zeigt die Dynamik des Saccharoseabbaus. Jede der Reaktionen kann im Rahmen der Reaktionsraten ausgeführt werden und trägt somit dem Saccharoseabbau und der Stärkebildung bei.

### 7.2.3 Stöchiometrische Analyse

Die stöchiometrische Analyse wird zunächst zur Überprüfung der Validität des Kartoffel-Modells untersucht. Dazu werden die *Dead-End* und *Gap-Finding* Analysen verwendet. Beide Analysen identifizieren den Metaboliten *AMP* als *Dead-End* Metaboliten. *AMP* wird in der Adenylatkinase Reaktion (*AdK*) zwar produziert, jedoch durch weitere Reaktionen nicht verbraucht. Somit kann *AdK* in dieser und auch in den vorherigen Analysen keinen Fluss tragen. Dieses Ergebnis bestätigt die Resultate der kinetischen und Petri-Netz Analysen.

Eine optimale Nutzung der Reaktionen im Kartoffel-Modell zur Bildung der maximalen Menge Stärke wird in der Flussbilanzanalyse (*FBA*) ermittelt. Die Zielfunktion Stärkebiosynthese (*StaSy*) ist in der quadratischen *FBA* maximiert.

Um konkrete Aussagen durch die *FBA* treffen zu können, werden verschiedene Reaktionskapazitäten auf die Reaktionsraten der kinetischen Analyseergebnisse im Fließgleichgewicht angepasst. Unter den Bedingungen eines konstanten Energieverbrauchs durch zelluläre Prozesse wird *ATPcons* auf einen Fließgleichgewichtswert von  $100 \mu M / FW / s$  gesetzt. Ebenso werden die Saccharose verbrauchenden Reaktionen Invertase ( $Inv = 0,16 \mu M / FW / s$ ) und Saccharose Synthase ( $SuSy = 4,89 \mu M / FW / s$ ) auf den Fließgleichgewichtswert eingeschränkt, sodass das bekannte Verhältnis von 3% zu 85% erreicht wird. Die Verwendung dieser drei Fließgleichgewichtswerte in der quadratischen *FBA* resultiert in einer Flussverteilung, die den kinetischen Fließgleichgewichtswerten entspricht, siehe Kantenskalierung in Abbildung 7.9.

Die Rate zur Bildung von Stärke entspricht mit  $6,42 \mu M / FW / s$  fast dem kinetischen Wert von  $6,22 \mu M / FW / s$ . Dieses Ergebnis zeigt, dass die *FBA* ohne detaillierte Beschreibungen der Reaktionsmechanismen die korrekte Rate zur Stärkebiosynthese ermittelt. Eine Ausnahme bilden die Reaktionen Saccharosephosphat Synthase (*SPS*)

und Saccharosephosphat Phosphatase (*SPP*), die in der optimalen Flussverteilung nicht berücksichtigt ( $0 \mu\text{M}/\text{FW}/\text{s}$ ) werden. Diese Reaktionen bilden zusammen mit der *SuSy* einen Saccharosezyklus. Über *SPS* und *SPP* würde Saccharose gebildet und durch *SuSy* abgebaut werden, ohne zur Stärkebiosynthese beizutragen. Diese Flussverteilung wäre keine optimale, sodass in der *FBA* die Reaktionen *SPS* und *SPP* keinen Fluss tragen.

Auch an dieser Stelle soll die Überexpression der Invertase die gesteigerte Stärkebiosynthese widerspiegeln. Wie in der vorherigen *FBA* werden die drei Reaktionskapazitäten auf die Fließgleichgewichtswerte gesetzt. Der Wert der *Inv* wird nun von  $0,16 \mu\text{M}/\text{FW}/\text{s}$  auf  $7,66 \mu\text{M}/\text{FW}/\text{s}$  erhöht, die quadratische *FBA* durchgeführt. Die Überexpression wird durch die geänderten Werte in der Flussverteilung (Kantenskalierungen) deutlich, siehe Abbildung 7.7.

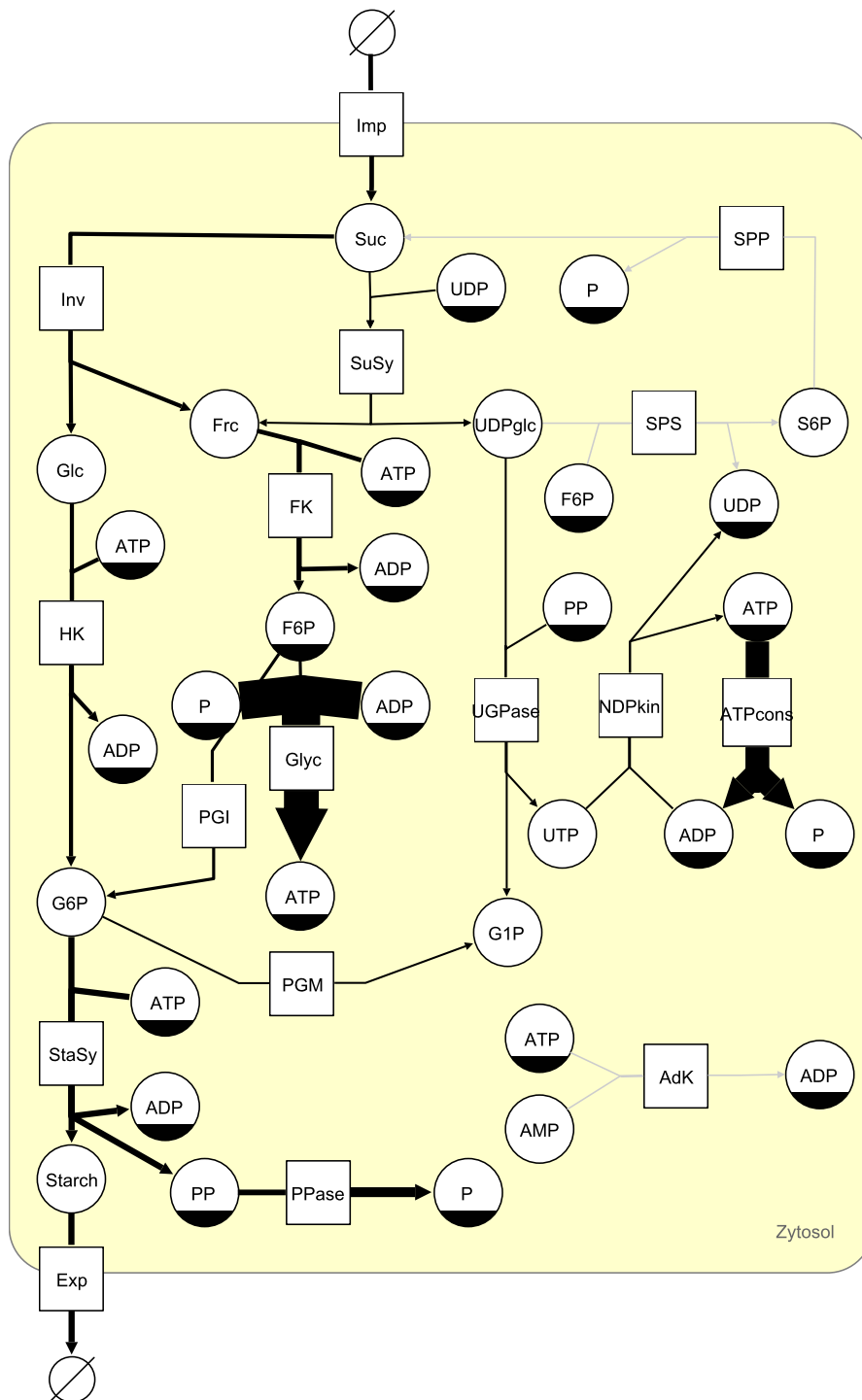
Die gesteigerte Stärkebiosynthese ist deutlich zu erkennen. Wie bereits in der kinetischen Analyse ermittelt, ist auch der Wert in der *FBA* im Vergleich zur vorherigen Analyse verdoppelt. Verglichen mit der kinetischen Analyse tragen die Reaktionen *SPS* und *SPP* weiterhin keinen Fluss. Die Reaktionen Uridindiphosphat Glukose Pyrophosphorylase (*UGPase*), Nucleosidediphosphat Kinases (*NDPkin*) und Phosphoglucomutase (*PGM*) unterscheiden sich in der Flussverteilung um den Faktor 1,6. Der Grund für die gleichen Flusswerte der Reaktionen *UGPase* und *NDPkin* liegt an der gemeinsamen Verwendung der Uridinnukleotide *UDPglc*, *UTP* und *UDP*. Die Biosynthese und der Abbau der Nukleotide ist nicht im Modell enthalten, sodass im Fließgleichgewicht die gleiche Menge produziert als auch verbraucht werden muss. Dieses Ergebnis entspricht genau dem der Petri-Netz P-Invarianten Analyse, die diese Nukleotide als eine P-Invariante berechnet [89]. Ausschlaggebend für den Flusswert ist die Einschränkung der Reaktionskapazität von *SuSy* auf  $1,38 \mu\text{M}/\text{FW}/\text{s}$ . Denn *SuSy* verbraucht *UDP* und bildet *UDPglc* mit genau dem Flusswert von *UGPase*, *NDPkin* und *PGM*. *PGM* verbraucht als einzige Reaktion den Metaboliten Glukose-1 Phosphat (*G1P*), welcher das Produkt der Reaktion *UGPase* ist. Damit entspricht der Flusswert von *PGM* denen der *UGPase*, *NDPkin* und *SuSy*. Die Kombination dieser Reaktionen wird in allen drei T-Invarianten, die den Saccharoseabbau über die *SuSy* beschreiben [89], repräsentiert.

Die stöchiometrische Analyse ermöglicht, ohne Berücksichtigung der Regulation und mit der Angabe von nur drei Werten, biologisch interessante Aussagen über das Kartoffel-Modell treffen zu können. Auch die Verifizierung wird mit der *Dead-End* und *Gap-Finding* Analyse ermöglicht.

#### 7.2.4 Topologische Analyse

Die topologische Analyse des Kartoffel-Modells erfolgt basierend auf dem entsprechenden Metabolitgraphen. Für die Erstellung einer Rangliste zur Identifikation wichtiger Metabolite des Kartoffel-Modells wird die Zentralitätsanalyse verwendet. Speziell findet die *Shortest-Path Betweenness (SPB)* Zentralitätsanalyse Anwendung. Diese berechnet eine Rangliste der Metabolite entsprechend ihrer Häufigkeit des Vorkommens auf den kürzesten Wegen zwischen je zwei Knoten. Das Analyseergebnis ist in Abbildung 7.8 im a) Kartoffel-Modell, b) Metabolitgraphen und c) in der Ergebnistabelle dargestellt.

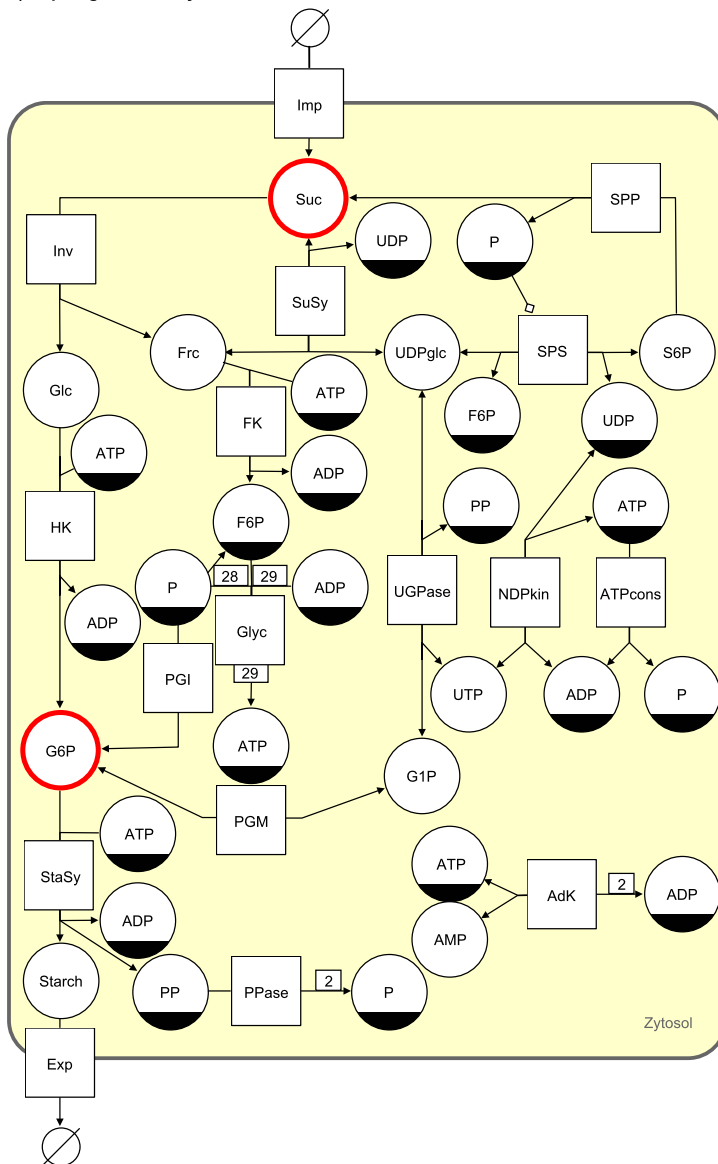
Die wichtigsten Metabolite des Kartoffel-Modells sind Glukose-6 Phosphat (*G6P*, *SPB* = 26,5), die energiereiche Verbindung *ADP* (*SPB* = 26), Saccharose (*Suc*, *SPB* = 20) und *ATP* (*SPB* = 17,3). Saccharose (*Suc*) und Glukose-6 Phosphat (*G6P*) sind sehr wichtige Metabolite für den Saccharoseabbauweg.



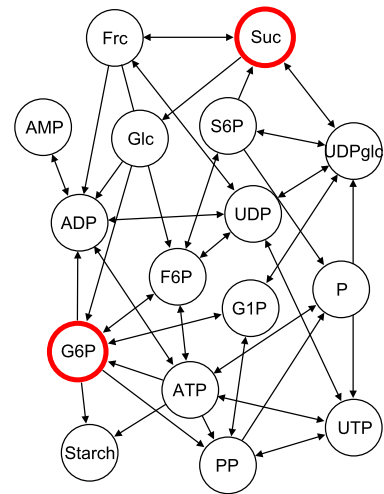
**Abbildung 7.7:** Die Flussbilanzanalyse (unter Invertaseüberexpression *Inv* und quadratischer Maximierung der Stärkebiosynthese *StaSy*) des Kartoffel-Modells resultiert in einer optimalen Flussverteilung. Die Kantenskalierung zeigt die Größe der Flusswerte an. (Eine hochaufgelöste Version der Abbildung ist auf der beiliegenden CD enthalten.)



a) Topologische Analyse des Kartoffel-Modells



b) Metabolitgraph



c) Ergebnistabelle der Zentralitätsanalyse

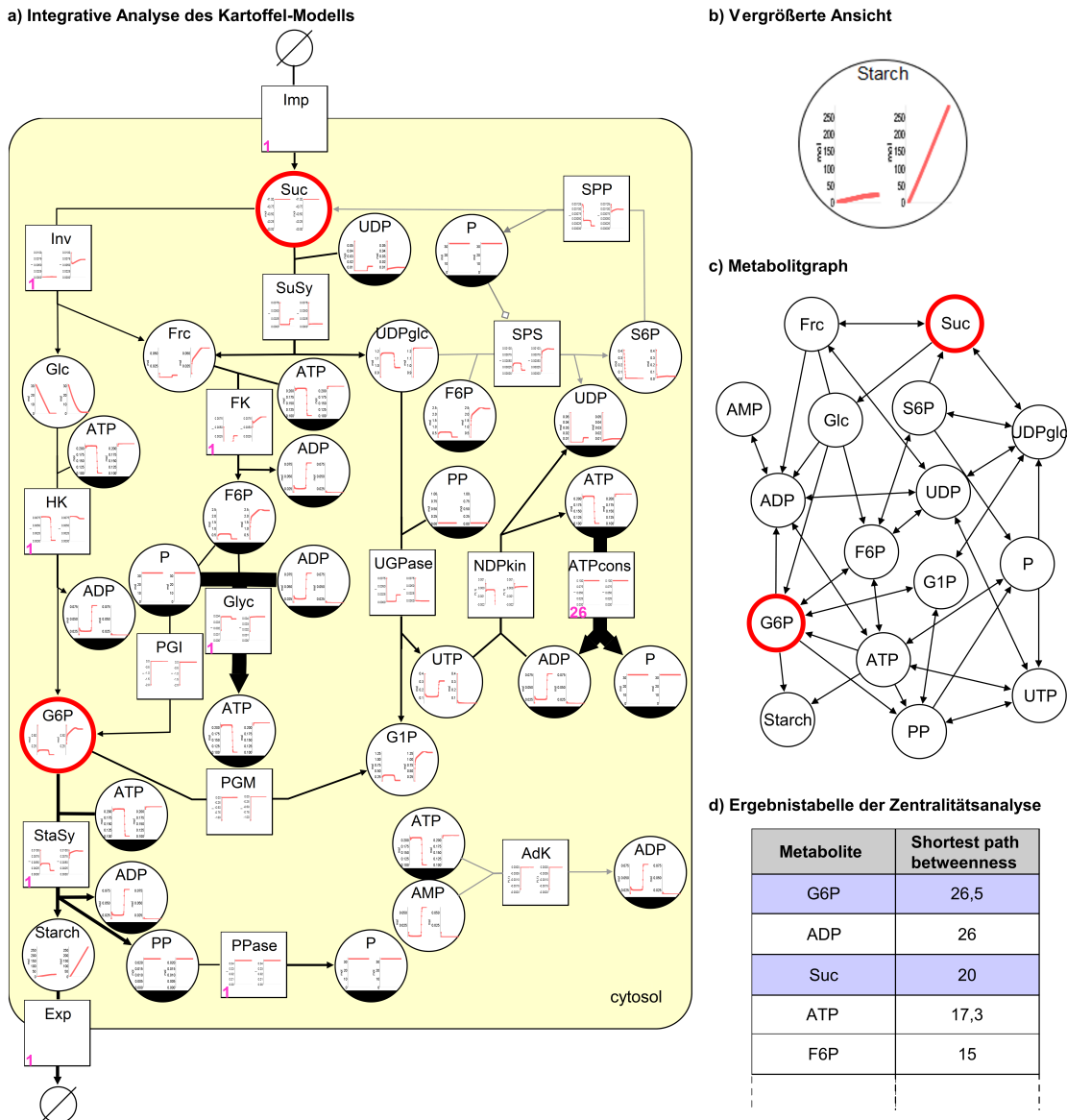
Metabolite	Shortest path betweenness
G6P	26,5
ADP	26
Suc	20
ATP	17,3
F6P	15

**Abbildung 7.8:** Die topologische *Shortest-Path Betweenness (SPB)* Zentralitätsanalyse, basierend auf dem b) Metabolitgraphen resultiert in einer Rangliste von Metaboliten, die in Form einer c) Ergebnistabelle repräsentiert wird. Die Auswahl der Metabolite *G6P* und *Suc* resultiert in einer Markierung (rote Umrandung) entsprechender Knoten im a) Kartoffel-Modell und b) Metabolitgraphen. (Eine hochaufgelöste Version der Abbildung ist auf der beiliegenden CD enthalten.)

Saccharose bildet den Start des Stoffwechselweges und wird von zwei Reaktionen gebildet (*Imp* und *SPP*) und von weiteren zwei Reaktionen verbraucht (*Inv* und *SuSy*). *G6P* ist die Vorstufe der Stärkebiosynthese und ist auch an vier Reaktionen beteiligt (*HK*, *PGI*, *PGM* und *StaSy*). *ATP* und *ADP* kommen sehr häufig in verschiedenen Reaktionen vor und verbinden reaktionsbedingt viele Metabolite des Metabolitgraphens miteinander. Würden diese Metabolite dem Modell entnommen werden, ist der Ablauf des Stoffwechsels nicht möglich. Wesentliche Reaktionen, z. B. die Stärkebiosynthese *StaSy* oder Glykolyse *Glyc* könnten nicht ablaufen.

### 7.2.5 Zusammenfassung

Das detailliert, quantitativ beschriebene Kartoffel-Modell wurde neben der kinetischen Analyse unter Verwendung zusätzlicher Modellierungsformalismen (Petri-Netz, stöchiometrische und topologische Analyse) aus weiteren Perspektiven analysiert. In Abbildung 7.9 ist zum Vergleich der unterschiedlichen Analyseergebnisse verschiedener Modellierungsformalismen die integrative Analyse zusammenfassend dargestellt.



**Abbildung 7.9:** Die integrative Analyse des Kartoffel-Modells beinhaltet: a) zwei kinetische Simulationen, die in jeweils zwei Diagrammen in den Knoten visualisiert sind (Zeitreihendiagramme links: initiale Parameter und rechts: Überexpression der Invertase *Inv*), b) ein vergrößerter Ausschnitt. Eine der berechneten Petri-Netz T-Invarianten (Schaltzähler in Magenta) stellt einen Hauptstoffwechselweg dar. Die Flussbilanzanalyse zur quadratischen Maximierung der Stärkebiosynthese resultiert in einer Flussverteilung, ähnlich der ersten kinetischen Simulation. Die topologische *Shortest-Path Betweenness (SPB)* Zentralitätsanalyse, basierend auf dem c) Metabolitgraphen, resultiert in d) essentiellen Metaboliten auf einem der Hauptstoffwechselwege (entsprechend der T-Invarianten) durch das Kartoffel-Modell. (Eine hochaufgelöste Version der Abbildung ist auf der beiliegenden CD enthalten.)

Diese integrative Analyse ermöglicht die Verwendung von Ergebnissen der kinetischen Analyse als Ausgangswerte der stochastischen Petri-Netz und der stöchiometrischen Flussbilanzanalyse. Es konnte gezeigt werden, dass neben der kinetischen Analyse weitere Analysen mit weniger detaillierten Daten interessante Aussagen treffen. Auch diese bestätigen und begründen andere Ergebnisse. Weiterhin ermöglicht die Integration von Ergebnissen in anderen Formalismen weiterführende Analysen. Die Interpretation der strukturellen und dynamischen Eigenschaften von Modellen wird durch sich ergänzende Analyseergebnisse unterstützt.

### 7.3 Integrative Anwendung metabolischer Modelle der Maispflanze

Im Sinne des Rekonstruktionsprozesses in der Systembiologie werden metabolische Modelle, abhängig von detailliertem Wissen über Reaktionsmechanismen, sukzessive aufgebaut. Ausgehend von qualitativen Modellen werden mit der stetigen Zunahme detaillierter quantitativer Experimentaldaten quantitative Modelle gebildet. Am Beispiel zweier metabolischer Modelle der Maispflanze wird dieser Prozess dargestellt.

Mais ist eine Energiepflanze mit einer hohen Biomassebildung und einem hohen Bioenergiegehalt. Als C4-Pflanze besitzt Mais im Gegensatz zu C3-Pflanzen u. a. eine effizientere Photosyntheserate und einen besseren Wasser-, Licht und Stickstoffnutzungsgrad [132]. Aus diesem Grund besitzt Mais ein deutlich höheres Biomassepotential als C3-Pflanzen. Durch die globale Klimaerwärmung ist zu erwarten, dass dieser Vorteil sogar noch zunimmt. Ein Ziel ist somit, die Systembiologie der Maispflanze durch metabolische Modellierung zu verstehen.

Bisher gibt es zwei genomweite stöchiometrische und einige kinetische Photosynthese-Modelle der Maispflanze. Die genomweiten stöchiometrischen Modelle sind C4GEM [35] und iRS1563 [133] und basieren auf genomischen Daten. Diese Modelle sind *Top-Down* erstellt wurden. Genomweite Modelle berücksichtigen die unterschiedliche Verwendung von Stoffwechselwegen in den verschiedenen Organen oder Geweben nicht. Im Oktober 2014 ist die zweite Generation des iRS1563 [148] veröffentlicht wurden. Es ist ein genomweites Modell des Maisblattes (keine Unterscheidung in Source- oder Sinkblatt), entstanden unter Verwendung von u. a. den Mais spezifischen Stoffwechselwegen aus dem MetaCrop Repository [140].

Anders hingegen die kinetischen Photosynthese-Modelle, die spezifisch für das Maisblatt *Bottom-Up* rekonstruiert wurden. Bereits vor 30 Jahren hat deren Entwicklung begonnen [13, 25]. Mit Weiterentwicklung der Technik wurden und werden diese Modelle verbessert sowie weitere Stoffwechselwege hinzugefügt [97, 98, 160]. Im April 2014 haben Wang und Koautoren [163] ein kinetisches Modell der C4-Photosynthese entwickelt. Es beinhaltet neben dem Calvin-Zyklus, der Stärkebiosynthese, Saccharosebiosynthese, dem C4-Zyklus, Photophosphorylierung und Kohlenstoffdioxid-Fixierung auch die Photorespiration und den Metabolitaustausch zwischen den beiden photosynthetisch aktiven Sourceblattzellen (Bündelscheiden- und Mesophyllzelle).

Die verschiedenen Organe und Gewebe der Maispflanze verwenden unterschiedliche Stoffwechselwege. Im Maiskorn findet hauptsächlich Stärkebiosynthese im Endospermgewebe statt, welche auch den größten Teil des Maiskorns bildet. In den Blättern (Sourceblätter) hingegen findet die Photosynthese zur Bereitstellung von Energie für die Bildung von Metaboliten statt. Aminosäuren und Saccharose werden

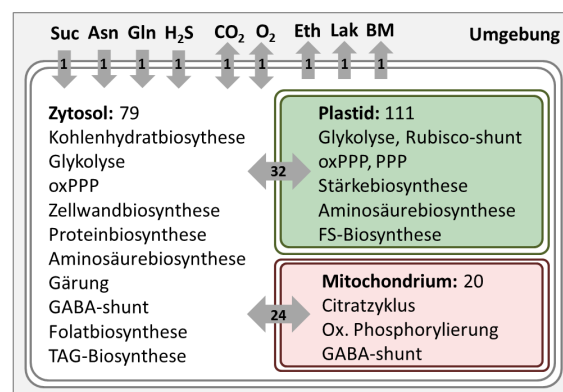
gebildet, exportiert und über den Stängel in die sich entwickelnden Organe und Gewebe, wie z. B. vom Maiskorn aufgenommen. Sourceblatt und Endosperm bilden somit eine interessante Basis, um den C4-Stoffwechsel im Sourceblatt und die Stärkebiosynthese im Endospermgewebe des Maiskorns zu untersuchen.

Grundlegend dafür ist zum einen die Rekonstruktion eines Biomasse produzierenden Modells des Maiskorns (Endosperm), um die metabolischen Prozesse in der Speicherphase zu untersuchen. Zum anderen die Rekonstruktion eines Sourceblatt-Modells zur Untersuchung des C4-Stoffwechsels. Die kuriierten Mais spezifischen Stoffwechselwege beider Modelle sind im MetaCrop Repository [140] hinterlegt.

### 7.3.1 Maiskorn-Modell

Das rekonstruierte Maiskorn-Modell ist ein stöchiometrisches Modell und bildet das größte Gewebe im Maiskorn (Endosperm) ab. Es ahmt die metabolische Kapazität während der Speicherphase nach. Mit dem Ziel, ein Biomasse produzierendes Modell zu erstellen, wurde zunächst die Biomassezusammensetzung im Endospermgewebe ermittelt. Aus Literaturangaben setzt sich die Biomasse aus 85% Kohlenhydraten (Zellwandbestandteile: 10% Zellulose, 4% Hemizellulose; 86% Stärke), 14% Proteinen (Aminosäuren), sowie 1% Lipiden zusammen [166]. Zur Biomassebildung werden einige Metabolite in das Endosperm importiert und andere exportiert. Aufgenommen werden die Aminosäuren Asparagin und Glutamin als Stickstoffquelle, Saccharose als Kohlenstoffquelle sowie Sauerstoff. Das Hauptabgabeprodukt ist Kohlenstoffdioxid. Mit diesen Angaben konnten die relevanten Stoffwechselwege erstellt werden. Diese bilden aus den aufgenommenen Metaboliten zunächst die Biomassevorstufen und anschließend die Biomasse.

Das Maiskorn-Modell besteht aus 41 Stoffwechselwegen, die auf drei Kompartimente verteilt sind. Insgesamt beinhaltet das Modell 250 Metabolite und 273 Reaktionen. Davon sind 225 enzymatische Reaktionen und 48 Transportprozesse. Zum Austausch der Metabolite mit der Umgebung und innerhalb der Zelle mit den Kompartimenten (Zytosol, Plastid und Mitochondrium) werden Transportprozesse verwendet, siehe Abbildung 7.10.

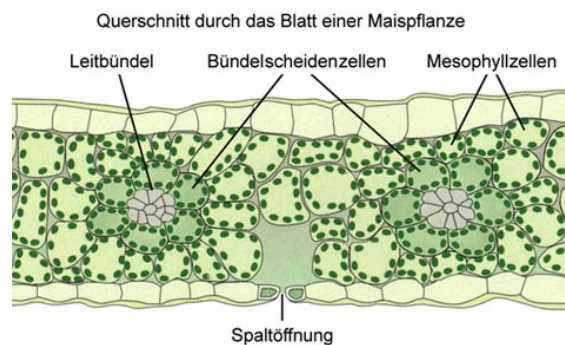


**Abbildung 7.10:** Schematische Darstellung des Maiskorn-Modells, bestehend aus 41 Stoffwechselwegen (250 Metabolite und 273 Reaktionen), die auf drei Kompartimente (Zytosol, Mitochondrium und Plastid) verteilt sind. Die Anzahl der biochemischen Reaktionen in den Kompartimenten ist angegeben. Transportreaktionen sind in Form grauer Pfeile dargestellt. Die Anzahl der Reaktionen ist in den Pfeilen wiedergegeben. Abkürzungen: Suc: Saccharose, Asn: Asparagin, Gln: Glutamin, Eth: Ethanol, Lak: Laktat, ox: oxidativ, PPP: Pentosephosphatweg, TAG: Triacylglycerol, FS: Fettsäure.

### 7.3.2 Maisblatt-Modell

Ein weiteres rekonstruiertes Modell der Maispflanze ist ein stöchiometrisches Modell. Es bildet den Stoffwechsel in einem vollständig entwickelten Maisblatt (Sourceblatt) ab. Dieses Modell ahmt die metabolische Kapazität während der Photosynthese nach. Es dient der Bildung von Saccharose und Aminosäuren, um die umgebenden Organe wie z. B. das Maiskorn mit den benötigten Metaboliten zu versorgen. Grundlage für den Stoffwechsel im Maisblatt ist die Aufnahme von Licht, Kohlenstoffdioxid als Kohlenstoffquelle, Nitrat als Stickstoffquelle sowie Sulfat als Schwefelquelle. Das Maisblatt-Modell besteht aus 29 Stoffwechselwegen, dem C4-Zyklus, der Photophosphorylierung, dem Calvin-Zyklus, Photorespiration, dem TCA-Zyklus und dem Stickstoff-, Sulfat- und Aminosäurestoffwechsel. Das Modell beinhaltet 310 Metabolite, die in 323 Reaktionen verstoffwechselt werden. Davon sind 205 biochemische Reaktionen.

Die *Kranzanatomie* des Maisblattes weist zwei verschiedene photosynthetisch aktive Zellen auf, die Mesophyll- und Bündelscheidenzelle, siehe Abbildung 7.11.



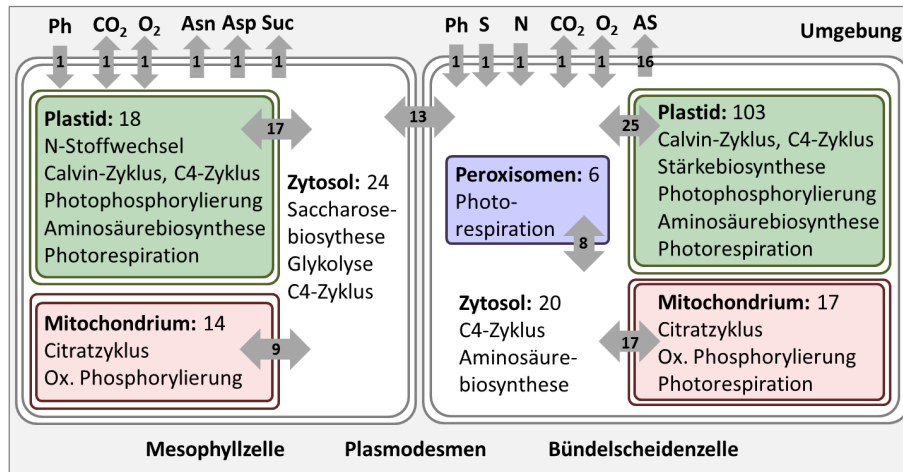
**Abbildung 7.11:** Querschnitt durch das Blatt einer Maispflanze. Die Kranzanatomie wird durch den Kranz der Zellen um die Leitbündel beschrieben, entnommen aus [157].

Der Vorteil dieser *Kranzanatomie* ist die Trennung zwischen Kohlenstoffdioxidfixierung in der Mesophyllzelle und der Kohlenhydratbiosynthese in der Bündelscheidenzellen. Die Mesophyllzelle pumpt  $CO_2$  in die Bündelscheidenzelle (zur Aufrechterhaltung der  $CO_2$ -Konzentration) sodass das Enzym Ribulosebiphosphat Carboxylase (*Rubisco*) die  $CO_2$ -Fixierung anstelle der Sauerstofffixierung durchführt. Somit wird die Photorespiration minimiert und eine höhere Kohlenhydratbiosynthese erreicht.

Die Stoffwechselwege finden im Modell in diesen beiden Zellen statt und werden mittels Plasmodesmen verbunden. Plasmodesmen ermöglichen gewissen Stoffwechselintermediaten den Austausch über die Zellgrenzen hinweg. Jede Zelle beinhaltet die Kompartimente Mitochondrium, Plastid und Zytosol. Die Bündelscheidenzelle besitzt zusätzlich ein Peroxisomen. 118 Transportprozesse verbinden die Stoffwechselwegintermediate über die Kompartiment- und Zellgrenzen hinweg, siehe Abbildung 7.12.

### 7.3.3 Topologische Analyse

Die topologische Analyse der Maispflanzen-Modelle wird basierend auf den zugehörigen Metabolit- und Reaktionsgraphen durchgeführt.

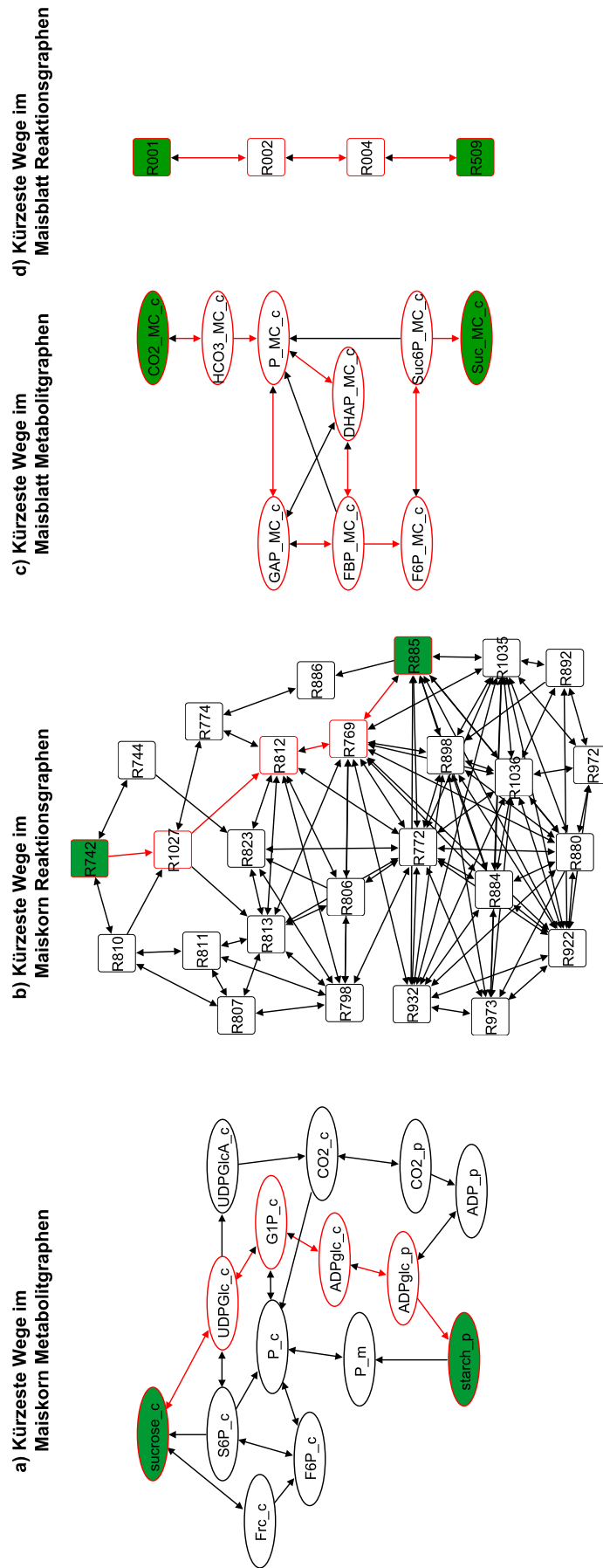


**Abbildung 7.12:** Schematische Darstellung des Maisblatt-Modells, bestehend aus zwei photosynthetisch aktiven Zellen, der Mesophyllzelle *MC* (links) und der Bündelscheidenzelle *BSC* (rechts) verbunden durch Plasmodesmen. Die *MC* besitzt die Kompartimente Zytosol, Mitochondrium und Plastid, die *BSC* zusätzlich das Peroxisomen. Das Modell beschreibt 29 Stoffwechselwege aus 310 Metaboliten und 323 Reaktionen. Die Anzahl der biochemischen Reaktionen ist in den Kompartimenten angegeben. Transportreaktionen sind in Form grauer Pfeile dargestellt. Die Anzahl der Reaktionen ist in den Pfeilen wiedergegeben. Abkürzungen: Ph: Photonen, Asn: Asparagin, Asp: Asparaginsäure, Suc: Saccharose, S: Sulfat, N: Stickstoff, AS: Aminosäuren, ox: oxidativ.

**Maiskorn-Modell** Der Hauptbestandteil der Biomasse im Maiskorn ist Stärke. Die Bildung von Stärke im Plastiden ( $starch_p$ ) aus der Kohlenstoffquelle Saccharose ( $sucrose_c$ ) im Zytosol beschreibt somit den Hauptweg der Biomassebildung. Deshalb wird die topologische Analyse des kürzesten Weges zwischen Saccharose und Stärke im Metabolit- und Reaktionsgraphen durchgeführt. Basierend auf den Metabolitgraphen ermittelt die Analyse des kürzesten Weges verschiedene kürzeste Wege. Einer verläuft über die vier Metabolite  $UDPGlc_c$ ,  $G1P_c$ ,  $ADPGlc_c$  im Zytosol und  $ADPGlc_p$  im Plastiden, siehe Abbildung 7.13 a). Diese Metabolite beschreiben die Zwischenprodukte der Stärkebiosynthese.

Im Reaktionsgraphen wird der kürzeste Weg zwischen den Reaktionen zur Saccharoseaufnahme ( $R742$ ) und Stärkebiosynthese ( $R885$ ) analysiert. Einer der kürzesten Wege verläuft über die Reaktionen Invertase ( $R1027$ ), Hexokinase ( $R812$ ) und dem ADP-Glukosetransporter ( $R769$ ), wobei der Kontakt zwischen den Reaktionen  $R812$  und  $R769$  nur über die Verbindung des Metaboliten  $ADP$  erreicht wird, siehe Abbildung 7.13 b).

Die Zentralitätsanalyse des Metabolitgraphen ermittelt eine Rangliste von Metaboliten. Phosphat ( $P$ ) im Zytosol und Glutamat ( $Glu$ ) im Plastiden sind in den Kategorien *Degree*, *Indegree*, *Outdegree* und *Shortest-Path Betweenness* (kurz *SPB*) unter den ersten zehn Metaboliten vertreten. Das Phosphat  $P$  im Zytosol erreicht die Zentralitätswerte (die Angaben von  $Glu$  in Klammern dahinter): *Degree* = 25 (26), *Indegree* = 14 (13), *Outdegree* = 11 (13) und *SPB* = 16.808 (4.581). Die höchsten Zentralitätswerte erreichen die Metabolite  $P$  im Plastiden in *Degree* (36) und *Indegree* (31),  $ATP$  im Plastiden in der Kategorie *Outdegree* (24) und das  $P$  im Zytosol in der *SPB* (16.808). Phosphate und energiereiche Verbindungen nehmen an sehr vielen Reaktionen teil und vernetzen den Metabolitgraphen stark.



**Abbildung 7.13:** Die Ergebnisse der topologischen Analyse der kürzesten Wege (rot hervorgehoben) zwischen zwei Knoten (grün hervorgehoben) der Mais-Modelle sind basierend auf den Metabolit- und Reaktionsgraphen dargestellt. Gezeigt werden die kürzesten Wege zwischen a) Saccharose (*sucrose\_c*) und Stärke (*starch\_p*) im Metabolitgraphen des Maiskorn-Modells, b) Saccharoseaufnahme (*R742*) und Stärkebiosynthese (*R885*) im Reaktionsgraphen des Maiskorn-Modells, c) Kohlenstoffdioxid (*CO2\_MC\_c*) und Saccharose (*Suc\_MC\_c*) im Metabolitgraphen des Maisblatt-Modells und d)  $CO_2$ -Diffusion (*R001*) und Saccharose-6 Phosphat Phosphatase (*R509*) im Reaktionsgraphen des Maisblatt-Modells.



Glutamat wird zur Biosynthese von Aminosäuren in den Plastiden verwendet. Die Bildung der Aminosäuren nimmt einen großen Teil des Maiskorn-Modells ein. Als Intermediat verbindet Glutamat somit die verschiedenen Aminosäurebiosynthesen.

Unter den zehn höchsten Zentralitätswerten, die aus der Analyse des Reaktionsgraphen stammen, sind mehr als die Hälfte in den vier Kategorien im Plastiden lokalisiert. Im Plastiden finden aufgrund der Aminosäurebiosynthese die meisten Reaktionen statt. In drei Kategorien (*Degree*, *Indegree*, *SPB*) sind die Reaktionen zur Fettsäurebiosynthese (*R1035*, *R1036* höchste Zentralität in *Indegree* = 54 und *SPB* = 10.952) und der Lysin Biosynthese (*R940*) im Plastiden vorhanden. Diese Reaktionen nutzen *ATP*, welches den Grund der hohen Vernetzung der Reaktionen untereinander realisiert. Die Reaktion *R940* verwendet *ATP* als Substrat und produziert die Metabolite *P* und *Glu*, die viele Reaktionen miteinander verbinden. Die Carbamoylphosphat Synthase (*R931*) gehört den Kategorien *Degree* (höchste Zentralität mit 71), *Indegree* und *Outdegree* mit hohen Zentralitätswerten an. Diese Reaktion findet im Plastiden statt und nutzt *ATP*, *CO<sub>2</sub>* und *Gln* um *ADP*, *P*, *Glu* und Carbamoylphosphat (*CP*) zu bilden. Im Vergleich zur Analyse der Metabolitgraphen werden die hochgradig vernetzten Metabolite *ATP*, *P* und *Glu* in dieser Reaktion verstoffwechselt.

**Maisblatt-Modell** Das Mais Sourceblatt nimmt Kohlenstoffdioxid auf und produziert Saccharose, die exportiert wird um weitere Pflanzenorgane, wie das Maiskorn, zu versorgen. Mit der topologischen Analyse soll der kürzeste Weg zwischen Kohlenstoffdioxid (*CO<sub>2</sub>*) und Saccharose (*Suc*) im Zytosol der Mesophyllzelle (*MC*) untersucht werden. Die Analyse des Metabolitgraphen ermittelt den kürzesten Weg über die Metabolite Hydrogencarbonat (*HCO<sub>3</sub>*), Phosphat *P*, Dihydroxyacetonphosphat (*DHAP*) oder Glycerinaldehyd-3 Phosphat (*GAP*) zu Fruktose-1,6 Bisphosphat (*FBP*), Fruktose-6 Phosphat (*F6P*) und Saccharose-6 Phosphat (*Suc6P*), siehe Abbildung 7.13 c). Grund für diesen kürzesten Weg ist das hochgradig vernetzte *P*, welches sehr viele Reaktionen miteinander verbindet und somit im Metabolitgraphen viele Metabolite untereinander verknüpft.

Weiterhin wird der kürzeste Weg im Reaktionsgraphen ermittelt. Der kürzeste Weg wird zwischen der Kohlenstoffdioxid Importreaktion (*R001*) und der Saccharose bildenden Reaktion Saccharose-6 Phosphat Phosphatase (*R509*) ermittelt. Es befinden sich die Carboanhydrase (*R002*) und die Phosphoenolpyruvatcarboxylase (*R004*) auf dem kürzesten Weg, siehe Abbildung 7.13 d). Auch hier ist das Phosphat *P* der verbindende Metabolit verschiedener Reaktionen, sodass Reaktion im Reaktionsgraphen stark untereinander verknüpft sind.

Die Betrachtung der Zentralitäten unterstützt die Ergebnisse der Analyse der kürzesten Wege. Das Phosphat *P* im Zytosol der Mesophyllzelle *MC* hat die folgenden Zentralitätswerte: *Degree* = 16, aufgeteilt in *Indegree* = 9 und *Outdegree* = 7. Auch die *Shortest-Path Betweenness* (kurz *SPB*) ist mit dem Wert 15.017 an vierter Stelle der Liste. Höhere Zentralitätswerte in diesen Kategorien erhalten das Phosphat *P* im Plastiden der Bündelscheidenzelle (*BSC*) mit den höchsten Werten in *Degree*, *Indegree* und *SPB* (*Degree* = 32, *Indegree* = 28, *Outdegree* = 4, *SPB* = 25.241), *ADP* und *ATP*. Die höchsten Werte werden in den *BSC* Plastiden erzielt, denn dort sind aufgrund der Aminosäurebiosynthese die meisten Stoffwechselwege lokalisiert.

Das Ergebnis der Zentralitätsanalyse des Reaktionsgraphen resultiert u. a. in drei Reaktionen des *BSC* Plastiden, die unter den zehn höchsten Zentralitätswerten der *Degree*, *Indegree*, *Outdegree* und *SPB* auftreten. Diese Reaktionen sind Triosephosphat Transporter (*R106*), Reaktion zur Erhaltungsenergie (*R712*) und Carbamoylphosphate Synthase



(*R911*). Diese Reaktionen verwenden *P* als Substrat oder Produkt. Die Reaktionen mit den höchsten und meisten (bis auf zwei Reaktionen) Zentralitätswerten aus den vier Kategorien sind in der *BSC* lokalisiert.

Die Zentralitätsanalyse der Metabolit- und Reaktionsgraphen beider Modelle der Maispflanze ermittelt Gemeinsamkeiten. Die Metabolite *P*, *ATP* und *Glu* haben in beiden Modellen die höchsten Zentralitätswerte der Kategorien *SPB*, *Degree* und *Outdegree* im Metabolitgraphen. Die Reaktion Carbamoylphosphate Synthase kommt in beiden Modellen mit hohen Zentralitätswerten in verschiedenen Kategorien vor. Grund dafür sind die übereinstimmenden Aminosäurebiosynthese Reaktionen beider Modelle, die sich wegen der Zugehörigkeit zum Zentralstoffwechsel nur in deren Lokalisation unterscheiden.

### 7.3.4 Stöchiometrische Analyse

Die beiden stöchiometrischen Modelle der Maispflanze wurden im Rekonstruktionsprozess mehrfach durch *Gap-Finding* und *Dead-End* Analysen untersucht und angepasst. Zur Vorhersage der optimalen Flussverteilung wird folgend die Flussbilanzanalyse *FBA* durchgeführt.

**Maiskorn-Modell** Im Maiskorn-Modell kann mittels quadratischer *FBA* durch Angabe der bekannten Saccharoseaufnahme von  $12,5 \mu\text{mol/gDW/h}$  (*R742*) aus der Literatur [2] *in silico* eine Wachstumsrate von  $0,004 \text{ h}^{-1}$  (*R998*) erzielt werden. Diese Wachstumsrate ( $0,0042 \text{ h}^{-1}$  [42]) entspricht der in der Literatur [2] angegebenen, siehe Kantenskalierung in der schematischen Abbildung 7.14.

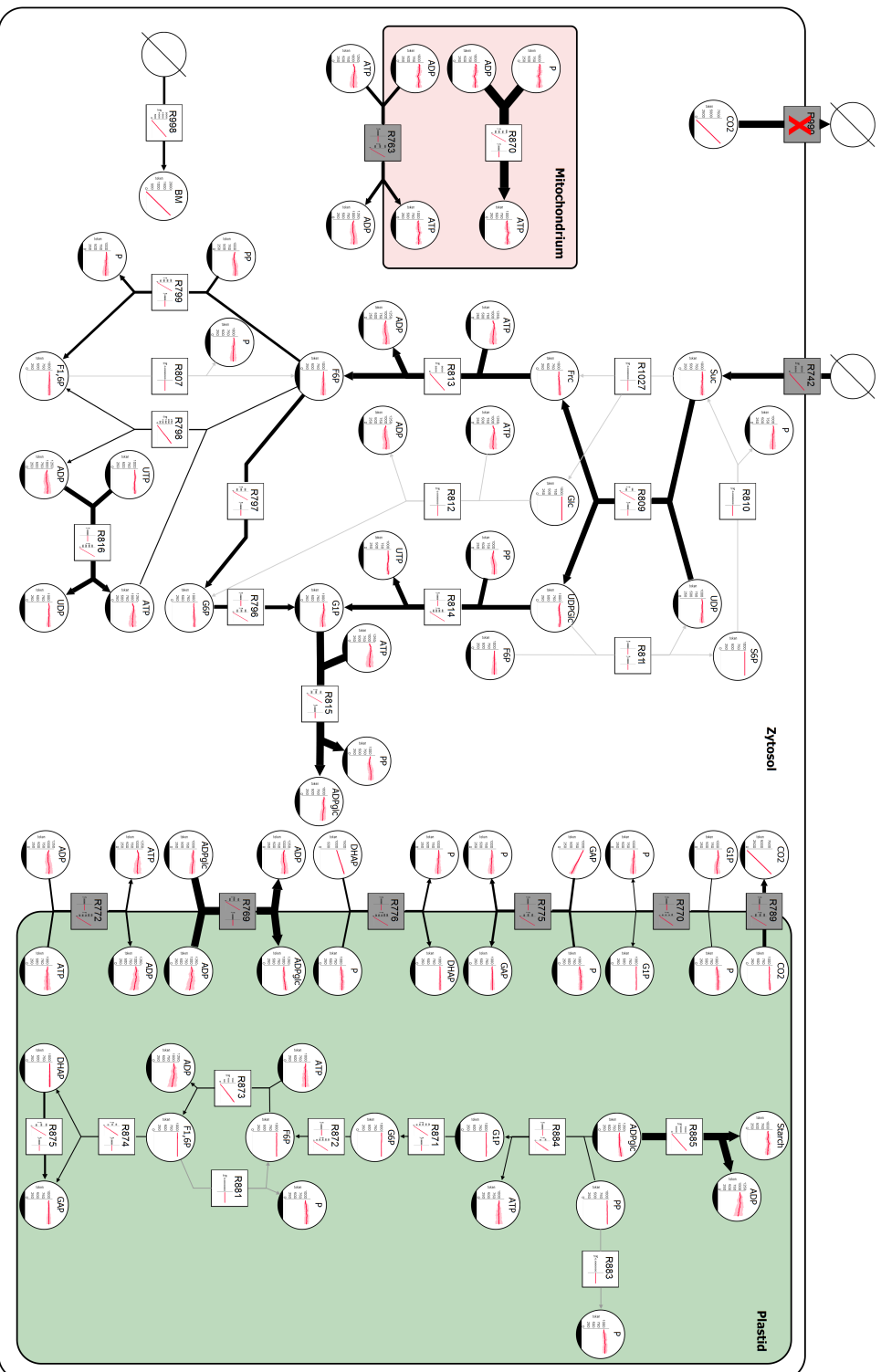
Der höchste Metabolitfluss verläuft von der Saccharoseaufnahme ( $12,5 \mu\text{M/gDW/h}$ ) über die Bildung von ADP-Glukose (*ADPglc*) im Zytosol (*R815* =  $17,2 \mu\text{M/gDW/h}$ ), welche zur Stärkebiosynthese in den Plastiden (*R885* =  $16,9 \mu\text{M/gDW/h}$ ) transportiert wird. Auch die oxidative Phosphorylierung im Mitochondrium trägt einen hohen Fluss ( $20,3 \mu\text{M/gDW/h}$ ) und generiert Energie (*ATP*) zur Biomassebildung.

In dem Artikel von Alonso und Koautoren [2] wurde der Zentralstoffwechsel des Endospermgewebes des sich entwickelnden Maiskorns mit metabolischer Flussanalyse (*MFA*) untersucht. Das entwickelte Modell berechnet die Flussverteilung für 44 biochemische Reaktionen. Davon sind 14 zusammengefasste Reaktionen, die Schlussfolgerungen zur Aminosäurebiosynthese, Glykolyse und dem TCA-Zyklus ermöglichen. Im Gegensatz zu diesem Modell erlaubt das in dieser Arbeit erstellte, detaillierte Maiskorn-Modell Flussverteilungen bis hin zu den Endprodukten der Aminosäurebiosynthese, verschiedene Zwischenschritte und Abzweigungen in Stoffwechselwegen zu untersuchen.

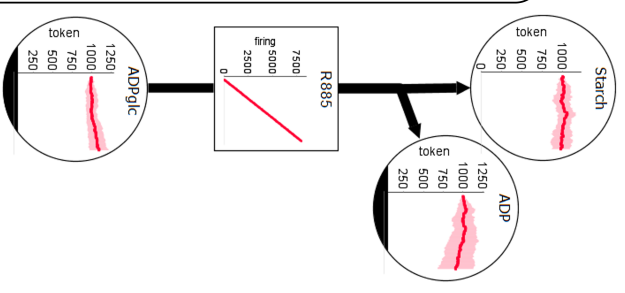
Im Vergleich mit der *MFA*-Flussverteilung erzielt die *FBA*-Flussverteilung mit der gleichen Saccharose-Aufnahmerate von  $12,5 \mu\text{M/gDW/h}$  im Wesentlichen einen unterschiedlichen Fluss zur Stärkebiosynthese (*MFA*  $7,3 \mu\text{M/gDW/h}$  zu *FBA*  $16,9 \mu\text{M/gDW/h}$ ). In der Arbeit von Alonso wird dieser geringe Stärkefluss mit Energie verschwendenden Substratzyklen erklärt. Es handelt sich um Triosephosphat- (z. B. *GAP*) und Hexosephosphat-Zyklen (z. B. *F6P*) zwischen Zytosol und Plastid, als auch Glukose- (*Glu*) und Hexosephosphat-Zyklen. Diese verbrauchen nach Alonso 47% Energie, die demnach zur Biomassebildung fehlt. Alonso postuliert, dass eine Unterbindung dieser Zyklen eine gesteigerte Biomassebildung erzielen könnte.

Die *FBA*-Flussverteilung zeigt genau diese beschriebene optimale Flussverteilung und verifiziert somit die Hypothese von Alonso.

a) Integrative Analyse des Maiskorn-Modells



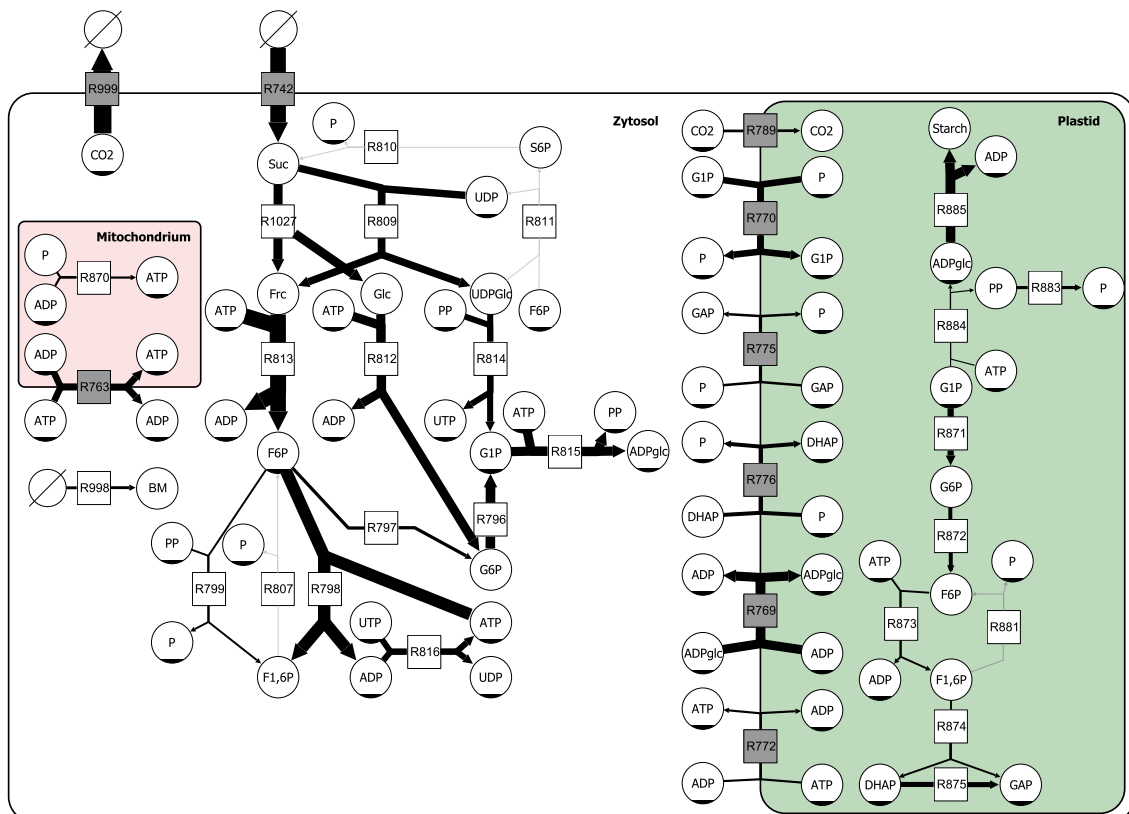
b) Vergrößerte Ansicht



**Abbildung 7.14:** Schematische Darstellung der integrativen Analyse des Maiskorn-Modells, bestehend aus a) der oxidativen Phosphorylierung im Mitochondrium, des Saccharoseabbaus im Zytoplast und der Stärkesynthese im Plastiden. Die Flussverteilung der quadratischen Flussbilanzanalyse (Kantenskalierung) und die Diagramme der stochastischen Petri-Netz Analyse sind dargestellt. Das rote Kreuz bedeutet, dass diese Reaktion von der Petri-Netz Analyse ausgeschlossen ist. b) Vergrößerte Ansicht der Stärkesynthese (R855 aus dem Plastiden). (Eine hochauflösende Version der Abbildung ist auf der beiliegenden CD enthalten.)

Unter Optimierung der Biomassebildung mit angegebener Saccharoseaufnahmerate finden diese energieverbrauchenden Substratzyklen nicht statt. Die Reaktionen, u. a. Fruktose-1,6 Bisphosphatase (*R807* im Zytosol und *R881* im Plastiden) und Invertase (*R1027*) werden nicht verwendet ( $0 \mu\text{M}/\text{gDW}/\text{h}$ ), sodass die Stärkebiosynthese um mehr als das Doppelte erhöht abläuft.

Zum besseren Vergleich der Flussverteilungen und Untersuchung der Substratzyklen wird eine zweite *FBA* durchgeführt. Dazu wird wie in [2] beschrieben die Saccharoseaufnahmerate von  $12,5 \mu\text{M}/\text{gDW}/\text{h}$  gesetzt und die Stärkebiosynthese auf  $7,3 \mu\text{M}/\text{gDW}/\text{h}$  eingeschränkt, siehe Abbildung 7.15.



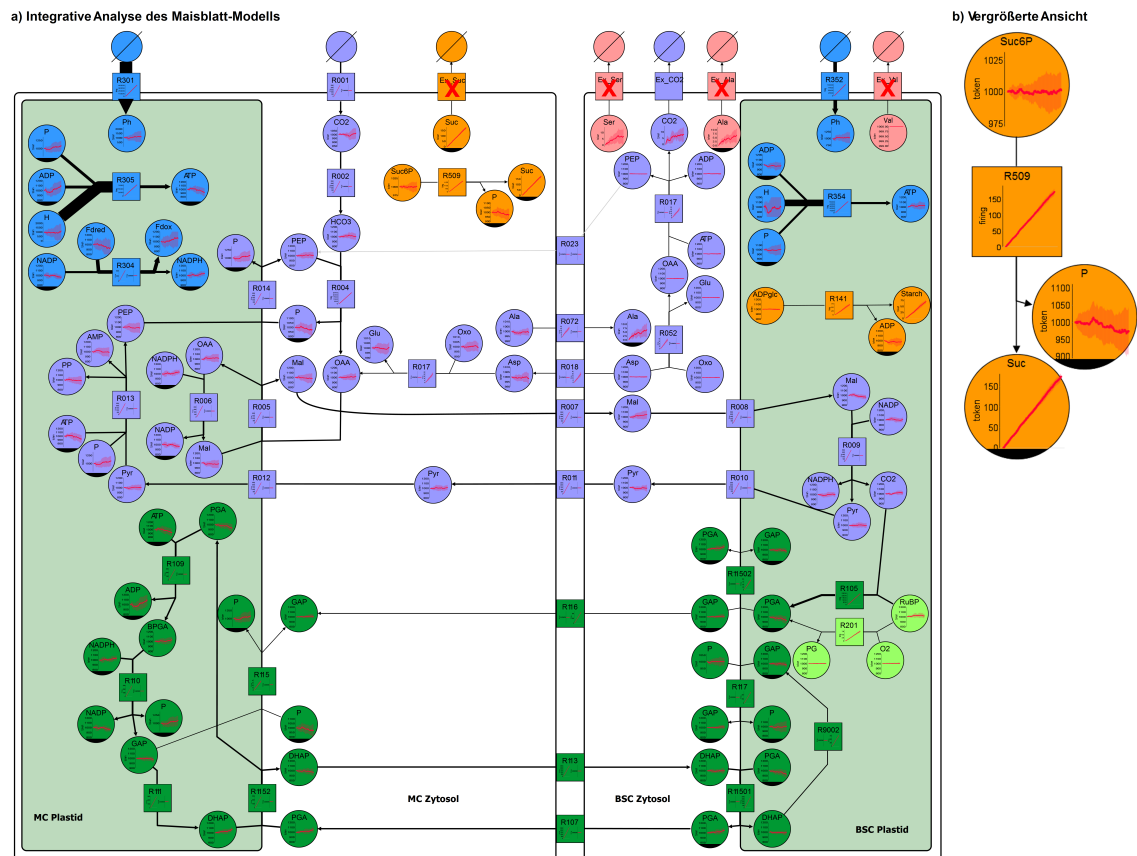
**Abbildung 7.15:** Schematische Darstellung der Flussbilanzanalyse des Maiskorn-Modells nach Alonso, bestehend aus der oxidativen Phosphorylierung im Mitochondrium, des Saccharoseabbaus im Zytosol und der Stärkebiosynthese im Plastiden. Die Flussverteilung der quadratischen Flussbilanzanalyse (Kantenskalierung) ist dargestellt.

Mit dem Ergebnis einer Wachstumsrate von  $0,0017 \text{ h}^{-1}$  (*R889* Flussrate von  $1,7454 \mu\text{M}/\text{gDW}/\text{h}$ ) können die von Alonso beschriebenen Substratzyklen beobachtet werden. Die Reaktionen Invertase (*R1027*) und Hexokinase (*R812*) laufen ab und generieren die energieverbrauchenden Substratzyklen. Die Biomassebildung ist um 42% geringer als im Vergleich zur optimalen Flussverteilung ohne Substratzyklen.

**Maisblatt-Modell** Das rekonstruierte Maisblatt-Modell wird unter Maximierung der Saccharoseexportrate quadratisch optimiert. Zur Untersuchung einer realistischen Flussverteilung wird das Modell in der Aufnahme von Licht (ambiante Lichteinstrahlung von  $5,86 \text{ mM}/\text{gFW}/\text{h}$ ) und Bildung von Aminosäuren nach Literaturangaben [72] eingeschränkt. Weiterhin berücksichtigt das Modell C4-Overcycling (10% der *PEPC*-Aktivität: *R017*

generiert  $CO_2$  Verlust:  $Ex\_CO2\_BC\_c$  in der BSC [4]) und die Erhaltungsenergie von 12% der Photosynthese [4]. Zusätzlich wird die Bildung von Stärke ( $R141$ ) und Saccharose ( $R509$ ) im Verhältnis 28%:72% [105] abgebildet. Die *RubisCo* Aktivität von Oxygenierung ( $R201$ ) zu Carboxylierung ( $R105$ ) wird auf das Verhältnis 3%:97% [36] eingestellt.

Die quadratische *FBA* resultiert in einer Flussverteilung, die die angegebenen Flüsse zur Aminosäurebildung erfüllt, jedoch den zu erwartenden Fluss zur Saccharosebildung ( $R509$  im Zytosol der Mesophyllzelle) nicht erzielt (nur 11%), siehe Kantenskalierung in der schematischen Abbildung 7.16.



**Abbildung 7.16:** Schematische Darstellung der integrativen Analyse des Maisblatt-Modells bestehend aus a) der Mesophyllzelle *MC* mit Plastiden und Zytosol (links), der Bündelscheidenzelle mit Zytosol und Plastid (rechts) und den Plasmodesmen zwischen den Zellen. Die Stoffwechselwege Photophosphorylierung (blau), C4- (violett), Calvin-Zyklus (dunkelgrün) und die erste Reaktion der Photorespiration (hellgrün) und die letzte Reaktion der Saccharose- (orange *MC*) und Stärkebiosynthese (orange *BSC*) sind dargestellt. Die Flussverteilung der quadratischen Flussbilanzanalyse (Kantenskalierung) und die Diagramme der stochastischen Petri-Netz Analyse werden gezeigt. Das rote Kreuz bedeutet, dass diese Reaktion von der Petri-Netz Analyse ausgeschlossen ist. b) Vergrößerte Ansicht der Saccharosebildung ( $R509$  aus dem Zytosol der Mesophyllzelle). (Eine hochauflösende Version der Abbildung ist auf der beiliegenden CD enthalten.)

Dennoch entspricht die übrige Flussverteilung den aus der Literatur bekannten Verhältnissen, wie im Folgenden erläutert. Der höchste Metabolitfluss verläuft in der Photophosphorylierung, gefolgt vom C4- und Calvin-Zyklus mit intensivem Transport zwischen den beiden Zellen und den Plastiden. Die Photophosphorylierung findet hauptsächlich in den Plastiden der *MC* statt. Es werden 60% Energie (*ATP*) und 100%

Reduktionsequivalente (*NADPH*) generiert. Im *BSC* Plastiden hingegen wird aufgrund des inaktiven Photosystems II [109] nur 40% *ATP* und kein *NADPH* erzeugt.

Der C4-Zyklus ermöglicht das im *BSC* Plastiden fehlende *NADPH* über den Transport und Abbau des zuvor in der *MC* gebildeten Malat (*Mal*) zu decken. 47% des gebildeten *NADPH* wird über den Malattransport im *BSC* Plastiden freigesetzt. Die Assimilation eines  $CO_2$  im Calvin-Zyklus bedarf zwei *NADPH* zur Reduktion von 3 Phosphoglycerinaldehyd (*PGA*). Es wird aber nur ein *NADPH* durch die Abgabe von  $CO_2$ , beim Abbau des *Mal* frei. Deshalb findet die Reduzierung des *PGA* hauptsächlich im *MC* Plastiden statt. Ein Teil des *PGA* wird deshalb von den *BSC* in die *MC* Plastiden transportiert. Aus *PGA* werden die Triosephosphate Glyzerinaldehyd-3 Phosphat (*GAP*) und Dihydroxyacetonphosphat (*DHAP*) produziert und zurück in die *BSC* Plastiden transportiert, um den Calvin-Zyklus zu beenden [165]. Die Transporte über die Plastidenmembranen hinweg werden durch den Triosephosphat/Phosphat Translokator (*TPT*) ermöglicht. Die Plasmodesmen ermöglichen die Diffusion der zu transportierenden Metabolite zwischen den Zellen.

Die C4-Photosynthese ist durch den koordinierten Transport von gebundenen Reduktionsequivalenten und das Ablaufen sich beeinflussender Stoffwechselwege in verschiedenen Zellen sehr effizient [163]. Im Modell werden somit hauptsächlich die Metabolite *PGA*, *DHAP*, *Mal* und Pyruvat (*Pyr*) zwischen den Zellen transportiert. Aber auch Alanin (*Ala*) und Aspartat (*Asp*) werden zum Austausch verwendet.

Pick und Koautoren [119] stellen eine Hypothese auf, dass neben *Mal* auch *Asp* über Plasmodesmen zur *BSC* diffundieren. Weiterhin wird behauptet, dass zur Regenerierung des Kohlenstoffakzeptors und dessen Transfer zur *MC* neben *Pyr* auch *Phosphoenolpyruvat* (*PEP*) und *Ala* genutzt werden könnten. In dem Modell konnte der Transport von *Asp* in die *BSC* und der Rücktransport durch *Ala* und *PEP* nicht erreicht werden. Das Gegenteil ist zu beobachten. *Asp* diffundiert von der *BSC* in die *MC* und *Ala* diffundiert aus der *MC* in die *BSC*. Die Diffusion von *PEP* findet gar nicht statt. Im Vergleich mit dem Diffusionsfluss von *Mal* ist das Verhältnis zu *Asp* 99% zu 1% und das Verhältnis von *Pyr* zu *Ala* 99% zu 1%.

Dieses Ergebnis stützt die Hypothese von Furbank [51], dass das Verhältnis von *Mal* zu *Asp* in den verschiedenen C4-Spezies variiert und der Transport von *Asp* in Bezug auf den indirekten Transport von *NADPH* durch *Mal* energetisch bevorzugt ist. Denn *Mal* versorgt die *BSC* mit dem fehlenden *NADPH*, welche durch den Transport von *Asp* nicht erfolgt. Das *Pyr* den Hauptfluss zurück zu *MC* beschreibt, erklärt Furbank [51] als höchstwahrscheinlich, da somit das Gleichgewicht der Aminogruppen zwischen beiden Zelltypen erhalten bleibt.

### 7.3.5 Petri-Netz Analyse

Zur Analyse des dynamischen Verhaltens der Modelle der Maispflanze wird die stochastische Petri-Netz Simulation durchgeführt. Die Angabe der Metabolitkonzentrationen zur Anfangsmarkierung des Petri-Netzes und Reaktionsraten sind Ausgangspunkt dieser Analyse.

**Maiskorn-Modell** Metabolitkonzentrationen zum Maiskorn-Modell wurden nicht erhoben und konnten auch nicht in der Literatur ermittelt werden. Auch die systematische Ermittlung der initialen Markierung konnte aufgrund der Größe des Modells nicht berechnet werden. Dazu sind verschiedene Kriterien zu berücksichtigen, u. a. dass jede P-Invariante mindestens eine Marke erhalten muss [55]. Eine Realisierung der Invariantenanalyse ist

für die Größe des Modells ohne Reduktion von z. B. linearen Stoffwechselwegen nicht möglich [89]. Deshalb wird für jeden Metaboliten (außer der Biomasse, deren Bildung beobachtet werden soll) eine initiale Metabolitkonzentration von  $1.000 \mu M$  gesetzt. Die Reaktionsraten zur stochastischen Analyse werden auf die Fließgleichgewichtsraten der zuvor berechneten *FBA* gesetzt.

Mit dem Ziel, die Biomassenakkumulation zu beobachten, wird die stochastische Analyse mit 200.000 Schritten und zehn Wiederholungen durchgeführt. Insgesamt können alle Reaktionen, die eine Rate gesetzt haben, schalten und Metabolite in unterschiedlichem Maße verstoffwechseln. Die Reaktionen mit den höchsten Flüssen im Fließgleichgewicht (Reaktionen der oxidativen Phosphorylierung *R870*, Saccharose-Aufnahme *R742* und Bildung der Stärkevorstufen und Stärke *R885*) schalten in der stochastischen Simulation sehr häufig, siehe Diagramme in der schematischen Abbildung 7.14. Der Anstieg der Kurven in den Diagrammen ist im Gegensatz zu Reaktionen, die kaum schalten, höher. Die Bildung der Biomasse kann mit der stetig ansteigenden Kurve im Diagramm bis auf eine Konzentration von  $1.940 \mu M$  beobachtet werden. Über die Schritte bleiben die Metabolitkonzentrationen mit kleineren Oszillationen zwischen den zehn Wiederholungen im Mittel konstant. Dabei oszillieren die Konzentrationen der energiereichen Metabolite (*ATP*, *ADP*, *NADPH*, *NADP*, *NADH*, *NAD*, *PP*, *P*) und Vorstufen der Stärke (*G1P*, *ADPglc*) in den Kompartimenten mehr als Metabolite, die an wenigen Reaktionen teilnehmen oder deren Reaktionsrate gering ist (z. B. Aminosäuren *Met*, *Ser*).

**Maisblatt-Modell** Zur stochastischen Analyse des Maisblatt-Modells werden die Metabolitkonzentrationen für die Anfangsmarkierung für alle Metabolite (außer die zu bildenden Metabolite Saccharose *Suc*, Stärke *Starch* und Aminosäuren) auf  $1.000 \mu M$  gesetzt. Im Gegensatz zum Maiskorn-Modell könnten einige Konzentrationen aus dem kinetischen Modell von Wang und Koautoren [163] verwendet werden, jedoch würden nur 26% (81 von 310) der Metabolite eine entsprechende Konzentration erhalten. Deshalb wird eine initiale Metabolitkonzentration von  $1.000 \mu M$  verwendet. Ähnlich verhält es sich mit den Reaktionsraten, die von 26 Reaktionen des kinetischen Modells verwendet werden könnten. Es handelt sich nur um 8% (26 von 323) aller Reaktionen im Maisblatt-Modell, sodass die Fließgleichgewichtsraten der *FBA*-Flussverteilung verwendet werden.

Die stochastische Analyse wird über 200.000 Schritte mit zehn Wiederholungen durchgeführt, um die Bildung der Saccharose (*Suc*), Stärke (*Starch*) und Aminosäuren zu beobachten. Jede Reaktion, die eine Fließgleichgewichtsraten besitzt, schaltet und verändert damit die Metabolitkonzentrationen zur Bildung der oben genannten Metabolite, siehe Diagramme in der schematischen Abbildung 7.16.

Nach 200.000 Schritten entstehen  $174 \mu M$  *Suc*,  $70 \mu M$  *Starch* und geringe Konzentrationen einiger Aminosäuren (Serin =  $7 \mu M$ , Glycin =  $9 \mu M$ , Glutamin =  $4 \mu M$  und Alanin =  $7 \mu M$ ). Aufgrund der stark unterschiedlichen Fließgleichgewichtsraten der Aminosäurebiosynthese zu allen anderen Raten schalten diese Reaktionen sehr selten oder gar nicht. Zum Beispiel wird die Aminosäure Valin (*Val*), mit einer Rate von  $0,0013 \mu M/gFW/h$  gebildet. Im Gegensatz dazu ist die Rate zur Bildung von *Suc* mit  $24,4 \mu M/gFW/h$  um einen Faktor von ca. 20.000 größer. Die Wahrscheinlichkeit, dass die Reaktionen der Aminosäurebiosynthese schalten, ist somit sehr gering und erklärt die geringen Aminosäurekonzentrationen.

Reaktionen der Lichtaufnahme und Photophosphorylierung schalten sehr häufig, wie in den stark ansteigenden Kurven der Diagramme zu erkennen ist. Die somit erzeugte Energie (*ATP*) und Reduktionäquivalente (*NADPH*) werden für z. B. die Reaktionen des C4- und

Calvin-Zyklus verwendet. Metabolitkonzentrationen der energiereichen Metabolite (*ATP*, *ADP*, *NADPH*, *NADP*, *PP*, *P*) oszillieren in den Plastiden der beiden Zellen innerhalb der zehn Wiederholungen. In den anderen Kompartimenten sind diese Oszillationen aufgrund der geringeren Reaktionsraten nicht zu beobachten. Im Mittel bleiben die Metabolitkonzentrationen konstant, sodass die Reaktionen weiterhin ablaufen können und *Suc*, *Starch* und Aminosäuren bilden.

### 7.3.6 Kinetische Analyse

Die kinetische Analyse der Maispflanzen-Modelle bedarf detaillierter Beschreibungen der Reaktionsmechanismen mit Parametern und initialen Metabolitkonzentrationen, die eine dynamische Analyse ermöglichen.

**Maiskorn-Modell** Im Fall des Maiskorn-Modells konnten diese quantitativen Experimentaldaten nicht erhoben werden. In der Literatur wurde bisher kein kinetisches Modell des Maiskorns erstellt, sodass die kinetische Analyse in der vorliegenden Arbeit nicht am Maiskorn-Modell angewendet werden kann.

**Maisblatt-Modell** Im Gegensatz zum Maiskorn gibt es einige kinetische Modelle der C4-Photosynthese [13, 25, 97, 98, 160, 163]. Das bisher umfassendste kinetische C4-Photosynthese-Modell ist jenes von Wang und Koautoren [163], welches im April 2014 veröffentlicht wurde. Dieses Modell berücksichtigt als Erstes die Lokalisation der Stoffwechselwege in den zwei photosynthetisch aktiven Zellen und damit auch den koordinierten Transport zwischen diesen.

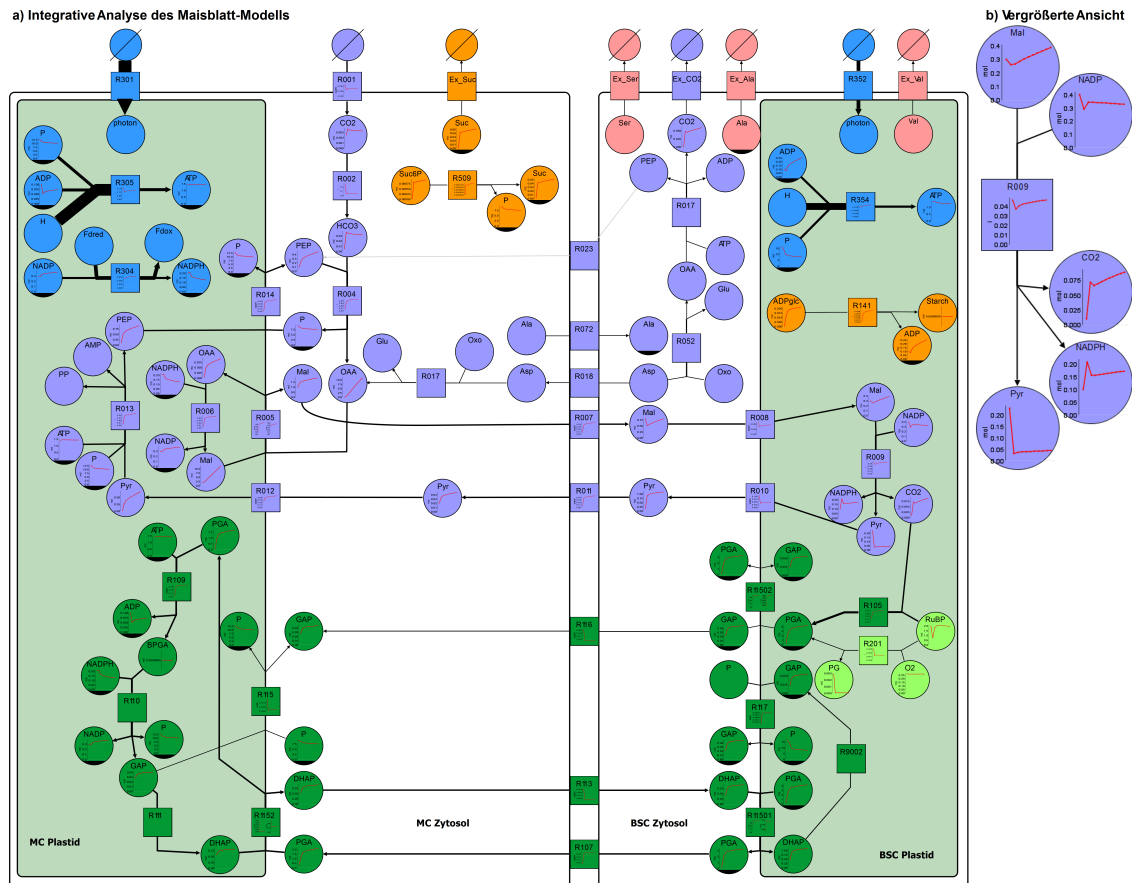
Das C4-Photosynthese-Modell besteht aus zwei Zellen, der Mesophyllzelle *MC* (mit den Kompartimenten Zytosol und Plastid) und der Bündelscheidenzelle *BSC* (Zytosol, Plastid und Peroxisomen). Es nehmen 91 Metabolite an 59 Reaktionen teil. Davon sind 11 Transportprozesse. Die Stoffwechselwege Calvin-Zyklus, Stärkebiosynthese, Saccharosebiosynthese, C4-Stoffwechsel, Photophosphorylierung, Kohlenstoffdioxid-Fixierung und Photorespiration werden im Modell abgebildet. Das Maisblatt-Modell enthält im Gegensatz zum C4-Photosynthese-Modell zusätzlich den TCA-Zyklus, die Stickstoff-, Sulfat- und Aminosäure-Stoffwechselwege.

Beim Vergleich beider Modelle stimmen nur 8% (26 von 323) Reaktionen und 26% (81 von 310) Metabolite überein. Das C4-Photosynthese-Modell beinhaltet zusammengefasste Reaktionen (z. B. *PGA Kinase* und *GAP Dehydrogenase*) und Metabolite, die sowohl in einer Menge als auch separat definiert werden. Zum Beispiel nimmt Fruktose-6 Phosphat zum einen in der Menge der Hexosephosphate *HexP* und zum anderen als separater Metabolit an Reaktionen teil (acht Metabolite werden in dieser Form verwendet). Mit der Berücksichtigung der verschiedenen Zellen und Kompartimente werden sowohl die Metabolite als auch die entsprechenden Mengen (Platzhalter für einen der Metabolite in der Menge) transportiert. Weiterhin wird die Lokalisation von Stoffwechselwegen nicht vollständig angewendet. Zum Beispiel erstreckt sich die Photorespiration über *MC* Plastid und *MC* Zytosol. Die meisten Reaktionen der Photorespiration findet im *BSC* Plastiden, *BSC* Zytosol, *BSC* Peroxisomen und *BSC* Mitochondrium statt. Im C4-Photosynthese-Modell werden nur der *BSC* Plastid und das *BSC* Peroxisomen verwendet.

Diese Vereinfachungen reduzieren die Komplexität der im C4-Photosynthese-Modell hinterlegten Kinetiken. Aufgrund der detaillierten strukturellen Beschreibung des

Maisblatt-Modells kann somit die Integration der Kinetik nicht direkt erfolgen. Entweder müsste das Maisblatt-Modell ebenso vereinfacht werden oder das C<sub>4</sub>-Photosynthese-Modell erweitert werden. Dennoch wird zum Vergleich der Fließgleichgewichtswerte beider Modelle die kinetische Analyse des C<sub>4</sub>-Photosynthese-Modells durchgeführt.

Zur Untersuchung des zeitlichen Ablaufs der biochemischen Reaktionen wird die Simulation mit der Adams-Bashforth Methode durchgeführt (Zeitspanne von 0 - 500.000, Schrittweite 50.000). Die Metabolitkonzentrationen und Reaktionsraten über die Zeit werden in Form von Zeitreihendiagrammen in das Modell integriert, siehe Diagramme in der schematischen Abbildung 7.17.



**Abbildung 7.17:** Schematische Darstellung der integrativen Analyse des Maisblatt-Modells bestehend aus a) der Mesophyllzelle *MC* mit Plastiden und Zytosol (links), der Bündelscheidenzelle mit Zytosol und Plastid (rechts) und den Plasmodesmen zwischen den Zellen. Die Stoffwechselwege Photophosphorylierung (blau), C<sub>4</sub>- (violett), Calvin-Zyklus (dunkelgrün), erste Reaktion der Photorespiration (hellgrün) und die letzte Reaktion der Saccharose- (orange *MC*) und Stärkebiosynthese (orange *MC*) sind dargestellt. Die Flussverteilung der quadratischen Flussbilanzanalyse (Kantenskalierung) und die Diagramme der kinetischen Analyse werden gezeigt. b) Vergrößerte Ansicht der Malat Dehydrogenase (*R009* aus dem Plastiden der Bündelscheidenzelle). (Eine hochauflösende Version der Abbildung ist auf der beiliegenden CD enthalten.)

In der kinetischen Analyse konvergieren die Metabolitkonzentrationen und Reaktionsraten nach einer Einschwingphase von etwa 400.000 s zum Fließgleichgewicht. Die höchsten Metabolitkonzentrationen erreichen die Metabolite Malat ( $Mal = 45.4$  mM) und Oxaloacetat ( $OAA = 41.3$  mM), im Zytosol der *MC*. Der höchste Fluss im Fließgleichgewicht wird durch die Photophosphorylierung (ATP Synthese Reaktionen *R305* und *R354*)



erreicht. Im Plastiden der *MC* wird ein Fluss von 2,1  $\text{mmol/gFW/h}$  *R305* und im Plastiden der *BSC* 1,6  $\text{mmol/gFW/h}$  *R354* erzielt. Im Verhältnis betrachtet, generiert die Photophosphorylierung im *MC* 57% Energie (*ATP*) und 100% Reduktionsequivalente (*NADPH*) zu 43% *ATP* und kein *NADPH* in der *BSC*. Denn 62% *NADPH* wird durch den Malattransport und die Malat Dehydrogenase Reaktion (*ME*, *R009*) im *BSC* Plastiden für den Calvin-Zyklus freigesetzt.

Verglichen mit dem Ergebnis der *FBA* des Maisblatt-Modells erreichen auch die Reaktionen der Photosphosphorylierung die höchsten Fließgleichgewichtsraten: *ATP* Synthese im *MC* Plastiden 1,1  $\text{mmol/gFW/h}$  *R305* und im *BSC* Plastiden 0,7  $\text{mmol/gFW/h}$  *R354*. Im Verhältnis sind es 60% *MC ATP* und 100% *MC NADPH R304*; 40% *BSC ATP* und kein *BSC NADPH*. Denn 47% *NADPH* wird durch den Malattransport und die *ME* Reaktion (*R009*) im *BSC* Plastiden für den Calvin-Zyklus freigesetzt.

Das Verhältnis der *RubisCo* Aktivität von Oxygenierung *R105* zu Carboxylierung *R201* ist 2%:98% im C4-Photosynthese-Modell und 3%:97% im Maisblatt-Modell. Der Plasmodementransport von *Mal* und *Pyr* erfolgt zu je 37%, der von *PGA* zu 25% und die Triosephosphate (*GAP* und *DHAP*) werden nur zu 0,4% ausgetauscht. Ein Austausch von Alanin und Aspartat wird im C4-Photosynthese-Modell nicht berücksichtigt.

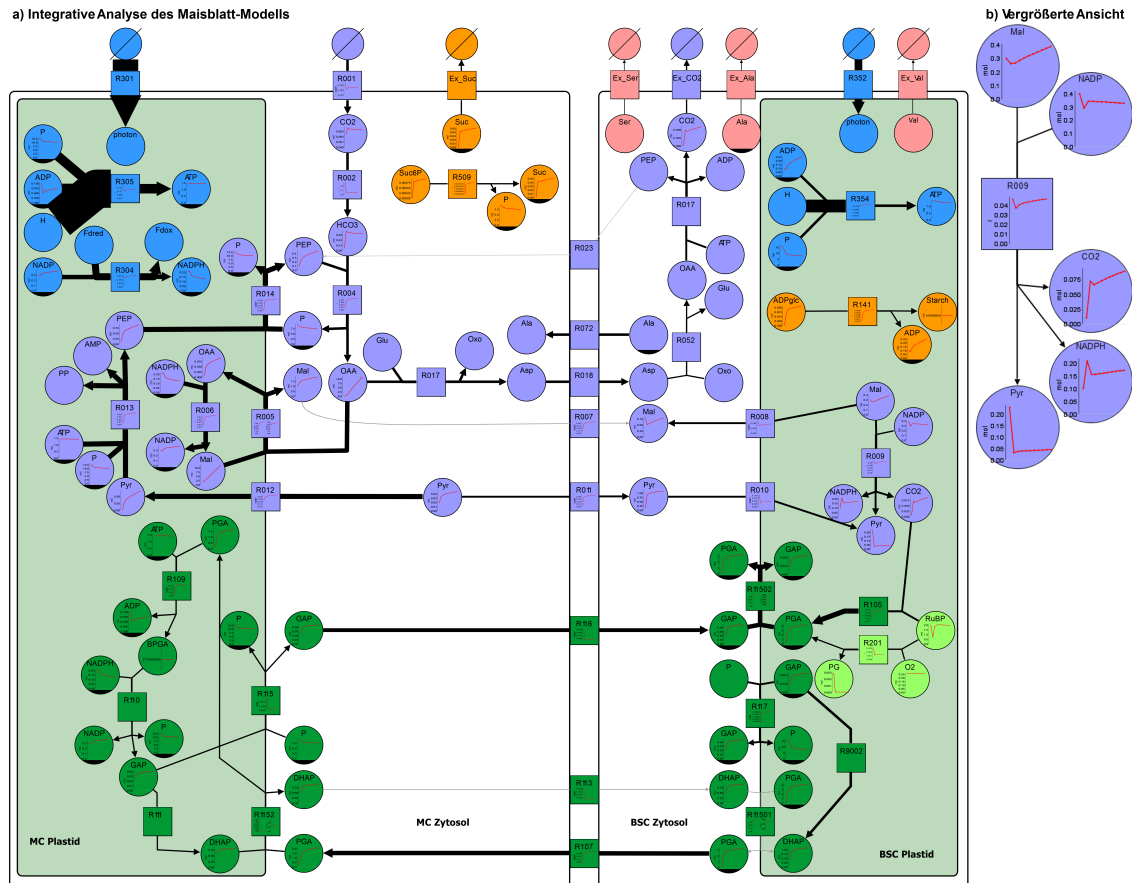
Die betrachteten Verhältnisse der kinetischen Analyse ähneln denen der *FBA* im Maisblatt-Modell. Die Fließgleichgewichtsraten direkt betrachtet (außer die übereinstimmenden Raten der Photorespiration), unterscheiden sich um den Faktor 2. Grundlegender Unterschied ist die lichtbedingte  $\text{CO}_2$  Aufnahme *R001*. Das Ergebnis der kinetischen Analyse ist eine  $\text{CO}_2$  Aufnahme von 0,64  $\text{mmol/gFW/h}$  ( $\text{CO}_2$  Aufnahme = *RubisCo* Carboxylierungsrate -  $0.5 \times$  *RubisCo* Oxygenierungsrate - mitochondriale Respirationsrate). Diese Aufnahme ist um einen Faktor 2 größer als die in der *FBA* erzielte Rate. Somit entsteht der Unterschied in den Flüssen.

Der größte Unterschied in den Analysen liegt in der Bildung von Stärke *R141* und Saccharose *R509*. Die kinetische Analyse berechnet ein Verhältnis von 98% Stärke zu 2% Saccharose. In der *FBA* ist das Verhältnis 28% Stärke zu 72% Saccharose entsprechend der Literaturangabe von Lunn und Hatch [105]. Die Biosynthese der Kohlenhydrate (Stärke und Saccharose) ist in der kinetische Analyse insgesamt um den Faktor 3 größer.

Mit dem Ziel, eine gesteigerte Saccharosebiosynthese in der *FBA* des Maisblatt-Modells zu erhalten, wird die *FBA* mit den kinetischen Fließgleichgewichtsraten wiederholt. Die maximal möglichen Aufnahmeleistungen von Licht und  $\text{CO}_2$  werden bei gleichbleibenden Bedingungen auf die Fließgleichgewichtswerte der kinetischen Analyse gesetzt. Das Ergebnis der *FBA* ist durch die erhöhten Aufnahmeleistungen dem Ergebnis der kinetischen Analyse ähnlicher, siehe Abbildung 7.18.

Durch die erhöhte Lichtaufnahme (*R301* und *R351*) wird mehr Energie (*R305* und *R354*) erzeugt und die erhöhte  $\text{CO}_2$  Aufnahme (*R001*) resultiert in einer höheren Kohlenhydratbiosynthese (Stärke *R141* und Saccharose *R509*, weiterhin im validen Verhältnis aus der Literatur). Die hohen Raten der kinetischen Analyse werden nicht erreicht, denn die *FBA* nutzt das  $\text{CO}_2$  zusätzlich zur Aminosäurebiosynthese. Die Steigerung der Saccharosebiosynthese kann auch mit Hilfe der Integration der kinetischen Daten nicht erzielt werden.

Generell zeigt die Integration der beiden Ansätze, dass die bekannten Verhältnisse der Fließgleichgewichtswerte (außer Saccharose zu Stärke) der *FBA* mit den Ergebnissen der kinetischen Analyse übereinstimmen. Die Vorhersagekraft der *FBA* des Maisblatt-Modells und die dynamische Berechnung einiger Reaktionen ermöglichen das Modellverhalten integriert zu untersuchen.



**Abbildung 7.18:** Schematische Darstellung der integrativen Analyse des Maisblatt-Modells bestehend aus a) der Mesophyllzelle *MC* mit Plastiden und Zytosol (links), der Bündelscheidenzelle mit Zytosol und Plastid (rechts) und den Plasmodesmen zwischen den Zellen. Die Stoffwechselwege Photophosphorylierung (blau), C<sub>4</sub>- (violett), Calvin-Zyklus (dunkelgrün), erste Reaktion der Photorespiration (hellgrün) und die letzte Reaktion der Saccharose- (orange *MC*) und Stärkebiosynthese (orange *MC*) sind dargestellt. Die Flussverteilung der quadratischen Flussbilanzanalyse (Kantenskalierung) mit integrierten kinetischen Werten und die Diagramme der kinetischen Analyse werden gezeigt. b) Vergrößerte Ansicht der Malat Dehydrogenase (*R009* aus dem Plastiden der Bündelscheidenzelle). (Eine hochauflösende Version der Abbildung ist auf der beiliegenden CD enthalten.)

### 7.3.7 Zusammenfassung

Die in dieser Arbeit entwickelten Mais-Modelle beschreiben die biochemischen Prozesse statisch und erlauben durch die Verwendung der integrativen Analyse dennoch quantitative Vorhersagen zu treffen. Das Maiskorn-Modell (273 Reaktionen, 250 Metabolite, 3 Kompartimente) und das Maisblatt-Modell (323 Reaktionen, 310 Metabolite, 2 Zellen, 7 Kompartimente) werden mit den verschiedenen Modellierungsfomalismen untersucht.

Die topologischen Analysen resultieren in Gemeinsamkeiten beider Modelle und begründen die dynamischen Analyseergebnisse. In den Flussbilanzanalysen konnten aufgestellte Hypothesen zum einen verifiziert und zum anderen widerlegt werden. Die entsprechenden Analyseergebnisse der Modelle erklärten die Sachverhalte im Detail. Die Integration dieser Analyseergebnisse als Ausgangswerte zur stochastischen Petri-Netz Analyse ermöglichten Einblicke in die Dynamik der Modelle. Eine detaillierte Beschreibung einiger Reaktionen des Maisblatt-Modells in einem kinetischen C<sub>4</sub>-Photosynthese-Modell

zeigt die Entstehung des Fließgleichgewichtes der *FBA*. Aufgrund der Vereinfachungen des kinetischen *C4*-Photosynthese-Modells konnte die Integration der Kinetik in das Maisblatt-Modell nicht erfolgen. Die Integration der berechneten Dynamik ermöglicht dennoch visuelle Analyse und Exploration im Kontext des Maisblatt-Modells und eine weiterführende *FBA* Analyse. Es werden keine Aussagen über die Kinetiken des Maiskorn-Modells getroffen, da es nicht kinetisch beschrieben ist.

Mit der Integration verschiedener Analyseergebnisse, zur weiterführenden Untersuchung in der *FBA* und der Petri-Netz Analyse, konnten die Mais-Modelle biologische Erkenntnisse hervorbringen. Insgesamt erlaubt die integrative Analyse somit, trotz fehlender detaillierter, quantitativer Beschreibungen von Modellen, dennoch Rückschlüsse auf die Verhaltens- und Funktionsweise der zugrunde liegenden biologischen Systeme zu ziehen. (Die verwendeten metabolischen Modelle sind auf der beiliegenden CD enthalten.)



## 8 Diskussion

In diesem Kapitel werden die Erkenntnisse und Resultate dieser Arbeit zusammengefasst und diskutiert. Abschließend wird ein Ausblick zukünftiger Anwendungen und Weiterentwicklungen gegeben. Das Vorgehen in diesem Kapitel ist in Abbildung 8.1 dargestellt.

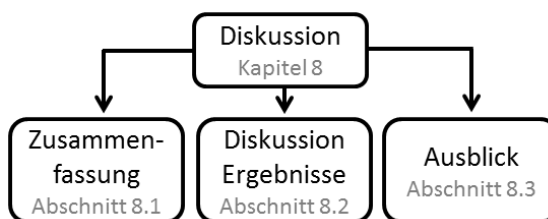


Abbildung 8.1: Struktur des Kapitels Diskussion.

### 8.1 Zusammenfassung

Die Charakterisierung von Verhaltens- und Funktionsweisen biologischer Systeme stellt aufgrund vielfältiger biochemischer Interaktionen eine große Herausforderung dar. Ansätze zur Modellierung und Analyse werden entwickelt, um die komplexen Mechanismen auf Systemebene zu verstehen. Verschiedene Modellierungsformalismen ermöglichen die Analyse metabolischer Modelle abhängig von der zu lösenden Fragestellung, dem biochemischen Wissen und der Verfügbarkeit experimenteller Daten.

Mit dem Ziel, metabolische Modelle aus mehreren Perspektiven analysieren zu können und somit das Potential verschiedener Formalismen integrativ zu nutzen, wurde eine Methodik entwickelt. Diese beinhaltet die Abstraktion, Formalisierung, Realisierung und Anwendung voneinander unabhängiger, etablierter Modellierungsformalismen zur metabolischen Modellierung. Zunächst wurden generelle Eigenschaften von Modellierungsformalismen untersucht, die ein breites Analysespektrum zulassen: kinetische, Petri-Netz (quantitativ und qualitativ), stöchiometrische (strukturell und optimierungsbasiert) und topologische Modellierung. In der Formalisierung wurde ein *abstrakter Graph* generalisiert, der zur Spezialisierung in unterschiedliche *spezifische Graphen* abgeleitet werden kann. Dieser generelle Ansatz ermöglicht die Integration bisher nicht kombinierbarer Modellierungsformalismen. Die integrative Analyse führt zum Schließen von Datenlücken in Modellen, sodass Analyseergebnisse im Kontext von Modellen in einer einheitlichen Visualisierung (*SBGN*) integrativ exploriert und interpretiert werden können. Schließlich werden somit neben strukturellen auch dynamische Eigenschaften des zugrunde liegenden biologischen Systems abgeleitet.

In der Evaluierung verschiedener Anwendungen und deren zugrunde liegender Methodik konnte kein integrierter Ansatz ermittelt werden, der die folgenden Anforderungen erfüllt: (1) Integration der Formalismen zur kinetischen, Petri-Netz, stöchiometrischen und topologischen Modellierung, (2) einheitliche Visualisierung (*SBGN*),

(3) Analyseergebnisse im Kontext des Modells und (4) Exploration und visuelle Analyse mittels Interaktionstechniken.

Aufgrund dessen wurde eine Methodik entwickelt, die eine Formalisierung zur Abstraktion des Metabolismus in einer einheitlichen und erweiterbaren Form vornimmt. Diese Formalisierung erlaubt metabolische Modelle zur Analyse in unterschiedlichen Modellierungsformalismen abzuleiten, ohne die Repräsentationsform zu ändern. Unzureichende Experimentaldaten, die möglicherweise zu fehlenden Ergebnissen der präziseren Analysen führen, werden durch Analyseergebnisse anderer Analysen ergänzt, die weniger detaillierte Daten benötigen. Zur Ausnutzung des Potentials verschiedener Modellierungsformalismen kann ein bereits detailliert, quantitativ beschriebenes metabolisches Modell unter Verwendung zusätzlicher Formalismen analysiert werden. Die sich ergänzenden Analyseergebnisse und auch die Integration dieser in anderen Formalismen, zur weiterführenden Analyse, unterstützen die Interpretation der strukturellen und dynamischen Eigenschaften des Modells. Unter Zuhilfenahme von Interaktionstechniken werden die Analyseergebnisse im Kontext der einheitlichen Darstellung des metabolischen Modells visualisiert. Die anschließende Exploration und visuelle Analyse führt zu einem umfassenden Verständnis für den Metabolismus des zugrunde liegenden biologischen Systems.

Realisiert ist die erarbeitete Methodik im *System Biology Metabolic Model Framework* (*SBM<sup>2</sup>-Framework*), welches die geforderten Anforderungen einer integrativen Anwendung umsetzt. Die integrative Analyse wurde erfolgreich an verschiedenen metabolischen Modellen von Kulturpflanzen in der Anwendung gezeigt. Im folgenden Abschnitt werden die erzielten Ergebnisse der vorliegenden Arbeit diskutiert.

## 8.2 Diskussion der Ergebnisse

Das zu Beginn gesetzte Ziel der vorliegenden Arbeit konnte mit der entwickelten Methodik erreicht werden. Mit der Abstraktion des Metabolismus wird eine Formalisierung durchgeführt, die zunächst die Definition eines *abstrakten Graphen* erzielt. Dieser erlaubt die Ableitung verschiedener *spezifischer Graphen*, deren Analysen, strukturelle und dynamische Eigenschaften metabolischer Modelle aufzeigen. Eine einheitliche Visualisierung des *abstrakten Graphen* ermöglicht die erhobenen Analyseergebnisse durch Datenzuordnungsfunktionen in das Modell zu integrieren. Aufgrund der erweiterbaren Formalisierung können neben den verwendeten Modellierungsformalismen weitere hinzugefügt werden. Dazu ist die Definition einer weiteren Ableitung des *abstrakten Graphen* vorzunehmen. Diese erlaubt den *spezifischen Graphen* des hinzugefügten Modellierungsformalismus zu erstellen, um Analysen durchführen zu können. Zur Integration der Analyseergebnisse und Visualisierung bedarf es keiner Anpassung. Es werden die definierten Datenintegrationsfunktionen und die Ableitung zur Visualisierung des *abstrakten Graphen* verwendet.

In der Realisierung wurde die entwickelte Methodik erfolgreich in Form des *SBM<sup>2</sup>-Framework* umgesetzt. Qualitative und quantitative metabolische Modelle können importiert, in *SBGN* visualisiert, mittels Modellierungsformalismen (kinetische, Petri-Netz, stöchiometrische und topologische Modellierung) analysiert und in die einheitliche Darstellung integriert werden. Bleiben aufgrund fehlender detaillierter Experimentaldaten einige Analyseergebnisse aus, verwendet der Anwender zum Ersatz der fehlenden Werte einen anderen Formalismus, der das vorhergehende Ergebnis ergänzt. Weiterhin profitiert der Anwender von bereits berechneten Analyseergebnissen, die als Parameter

in einen anderen Formalismus eingesetzt werden können. Diese integrative Analyse nutzt das Potential der verschiedenen Modellierungsformalismen aus und ermöglicht neue Erkenntnisse des modellierten biologischen Systems hervorzubringen. Aufgrund des modularen Aufbaus des Frameworks können über den Addon Mechanismus von VANTED weitere Modellierungsformalismen hinzugefügt werden.

Die Anwendung des *SBM<sup>2</sup>-Frameworks* zur integrativen Analyse struktureller und dynamischer Eigenschaften metabolischer Modelle liefert neue Erkenntnissen. Verschiedene spezifische qualitative und quantitative Modelle wurden verwendet und konnten unter anderem, die in der Literatur beschriebenen Analyseergebnisse erzielen. Die integrative Analyse des Kartoffel-Modells (mittlere Größe: 15 Reaktionen, 17 Metabolite, 1 Kompartiment) ermöglichte nicht nur bekannte Resultate zu reproduzieren. Unter Verwendung der stochastischen Petri-Netz Analyse und stöchiometrischen Flussbilanzanalyse (*FBA*) konnten trotz Verwendung weniger detaillierter Experimentaldaten dynamisches Verhalten beobachtet und quantitative Vorhersagen getroffen werden. Die Ergebnisse der kinetischen Analysen konnten mit den Resultaten der weiteren Analysen ergänzt und bestätigt werden.

Die Anwendung zweier großer metabolischer Modelle der Maispflanze bestätigen und widerlegen bisher in der Literatur aufgestellte Hypothesen. Die Modelle vom Maiskorn (273 Reaktionen, 250 Metabolite, 3 Kompartimente) und Maisblatt (323 Reaktionen, 310 Metabolite, 2 Zellen, 7 Kompartimente) sind mit den verschiedenen Modellierungsformalismen integrativ untersucht wurden. Ursprünglich wurden diese Modelle zur *FBA* entwickelt und enthalten nur eine statische Beschreibung der Reaktionsmechanismen. Die Analyseergebnisse der *FBA* bestätigen und widerlegen in der Literatur aufgestellte Hypothesen und liefern die Erklärung verschiedener Annahmen. Diese Annahmen konnten mit den Modellen im Detail nachvollzogen werden.

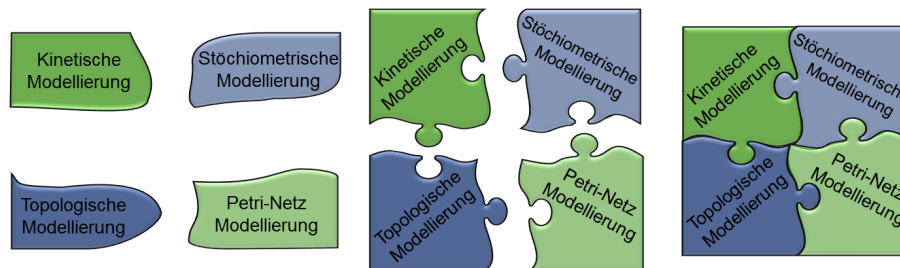
Mit der Integration der Analyseergebnisse, als Ausgangswerte zur stochastischen Petri-Netz Analyse, konnte neben dem Fließgleichgewicht auch die Dynamik der Modelle gezeigt werden. Im Fall des Maisblatt-Modells sollte die Bildung der Saccharose, Stärke und Aminosäuren beobachtet werden. In der *FBA* ist die Bildung dieser Metabolite im Fließgleichgewicht möglich. Die stochastische Petri-Netz Analyse hingegen generiert hauptsächlich die Metabolite Saccharose und Stärke, jedoch nur geringe Mengen einiger Aminosäuren. Der Unterschied zwischen den verwendeten Reaktionsraten im Fließgleichgewicht ist zu groß. Es werden Raten in verschiedenen Größenordnungen zugleich verwendet. Die Rate der Saccharosebiosynthese ist mit  $24,4 \mu\text{M}/\text{gFW}/\text{h}$  viel Größer als die Rate der Valinbiosynthese mit  $0,0013 \mu\text{M}/\text{gFW}/\text{h}$ . Die Wahrscheinlichkeit, dass eine Reaktion mit einer niedrigen Rate schaltet, ist sehr gering. Die stochastische Petri-Netz Analyse ist bereits über 200.000 Schritte und zehn Zyklen durchgeführt wurden. Dabei schalteten auch Reaktionen mit sehr kleinen Raten. Die Wiederholung der stochastischen Petri-Netz Analyse mit z. B. 1.000.000 Schritten würde die Anzahl des Schaltens von Reaktionen mit geringen Raten erhöhen. Reaktionen mit hohen Raten schalten auch häufiger. Das Verhältnis zwischen der Saccharosebiosyntheserate zu den Aminosäurebiosyntheseraten bleibt erhalten.

Eine Interpretation der Ergebnisse lässt auf nicht korrekt Verhältnisse schließen, obwohl die verwendeten Experimentaldaten zur Modellierung aus der Literatur stammen. In der *FBA* konnte die zu erzielende Saccharoseexportrate nicht erreicht werden, wenn auch die Verhältnisse und Raten aus der Literatur mit dem Modell übereinstimmen. An dieser Stelle sollten Folgeexperimente zur Bestimmung von Saccharose- und Aminosäureexportraten

durchgeführt werden, um weitere Referenzwerte zu ermitteln. Somit könnten die Aussagen des Modells besser verifiziert werden.

Detaillierte Beschreibungen einiger Reaktionen des Maisblatt-Modells in einem kinetischen C4-Photosynthese-Modell zeigt die Entstehung des Fließgleichgewichtes der *FBA*. Aufgrund der Vereinfachungen des kinetischen C4-Photosynthese-Modells konnte die Integration der Kinetik in das Maisblatt-Modell nicht erfolgen. Die Integration der berechneten Dynamik ermöglicht dennoch die Exploration und visuelle Analyse im Kontext des Maisblatt-Modells. Leider war es nicht möglich die Aminosäurebiosynthese kinetisch zu analysieren, da das kinetische Modell diese nicht beinhaltet. Es werden keine Aussagen über die Kinetik des Maiskorn-Modells getroffen, da es nicht kinetisch beschrieben ist. Im Vergleich beider Mais-Modelle resultieren die topologischen Analysen in Gemeinsamkeiten der Modelle und begründen dynamische Analyseergebnisse. Die integrative Analyse erlaubt somit, trotz fehlender detaillierter quantitativer Beschreibungen von Modellen, dennoch Rückschlüsse auf die Verhaltens- und Funktionsweise der zugrunde liegenden biologischen Systeme zu ziehen.

Das Ergebnis der vorliegenden Arbeit wird in der Abbildung 8.2 schematisch dargestellt.



**Abbildung 8.2:** Die vier voneinander unabhängigen Modellierungsfomalismen (links) werden in der erarbeiteten Methodik integriert (Mitte) und in der Realisierung miteinander verbunden (rechts).

Die integrative Analyse der Struktur und Dynamik metabolischer Modelle verbindet voneinander unabhängige Modellierungsfomalismen. Das Verbinden der in der Methodik integrierten Formalismen wird in der Realisierung (*SBM<sup>2</sup>-Framework*) erzielt. Die berechneten Ergebnisse der erfolgreichen Anwendung verschiedener metabolischer Modelle zeigt den Nutzen der vorgestellten Arbeit.

### 8.3 Ausblick

In der vorliegenden Arbeit werden metabolische Modelle vielfältig analysiert, um ein tieferes Verständnis der Verhaltens- und Funktionsweise des zugrunde liegenden biologischen Systems zu erhalten. Die in Kapitel 7 Abschnitt 7.3 rekonstruierten und integriert analysierten Modelle der Maispflanze könnten zukünftig nach dem Prinzip von Grafahrend-Belau und Kollegen [61] zu einem metabolischen Gesamtpflanzen-Modell verbunden werden. So könnten die Prozesse einer gesteigerten Biomasseproduktion in der Energiepflanze Mais erforscht werden.

Im Hinblick auf die *SBM<sup>2</sup>-Framework* Anwendung sollte insbesondere hinsichtlich der Performance stochastischer Petri-Netz und kinetischer Analysen optimiert werden. Die Anwendung großer Modelle (> 500 Knoten) verbraucht zu viel Speicher und ist zeitintensiv. Die Erweiterung des Frameworks zur Unterstützung einer Editorfunktionalität würde neben



der vorhanden Funktionalität der Parameteranpassung die Rekonstruktion metabolischer Modelle ermöglichen.

Des Weiteren werden in der Literatur, neben der Verwendung von metabolischen Modellen, auch Genregulations- und Signaltransduktionsmodelle in topologischer, stöchiometrischer, Petri-Netz und kinetischer Modellierung beschrieben. Eine Erweiterung der Methodik und der Realisierung zur Analyse dieser zusätzlichen Modelltypen könnte weitere Einblicke in die komplexe Regulation des Metabolismus ermöglichen. Die topologische Modellierung wird von Koschützki [91] unter Verwendung aller drei Modelltypen (metabolisch, Genregulations- und Signaltransduktionsmodelle) diskutiert und für metabolische Modelle und Genregulationsmodelle angewendet. Die stöchiometrische Modellierung ist für metabolische Modelle entwickelt wurden [115], wird jedoch zur Integration der drei Modelltypen in der Modellierung der gesamten Zelle verwendet [27, 28, 101, 106, 146]. Petri-Netz Modellierung wird von Claudine und Koautoren [22] sowie von Machado und Koautoren [106] für die drei Modelltypen erläutert. Baldan und Koautoren [8] beschreiben die Anwendung von Petri-Netzen für metabolische Modelle, De Jong [34] für Genregulationsmodelle und Gibert und Koautoren [56] für Signaltransduktionsmodelle. Die kinetische Modellierung ist für alle Modelltypen möglich, wie die Anwendung von Machado [106] zeigt.

Die entwickelte Methodik kann um Genregulations- und Signaltransduktionsmodelle erweitert werden. Der *abstrakte Graph* würde dazu um weitere Knoten- und Kantenattribute erweitert werden. Es könnten weitere Ableitungen hinzugefügt werden, um die *spezifischen Graphen* zur Analyse zu generieren. Die Visualisierung nutzt *SBGN-PD* für die metabolischen Modelle. Die Teilsprache *SBGN-AF* würde zur Visualisierung der Genregulationsmodelle und *SBGN-ER* für die Signaltransduktionsmodelle verwendet werden.

Unter diesen Voraussetzungen ist neben der Entwicklung eines metabolischen Gesamtpflanzen-Modells auch die Integration von Modellen der Signaltransduktion, Genregulation und des Metabolismus zukünftig möglich. Ziel ist es, das Verhalten von Organismen mit allen Interaktionen zwischen den verschiedenen Ebenen zu analysieren. Karr und Koautoren [80, 81] zeigen dieses Vorgehen bereits am Beispiel eines menschlichen Krankheitserregers. Wege, dieses Ziel zu erreichen sind auch in Reviews [10, 57, 122, 155, 159] beschrieben.



---

# Literaturverzeichnis

- [1] ADIAMAH, D. A. und J.-M. SCHWARTZ: *Construction of a Genome-Scale Kinetic Model of Mycobacterium Tuberculosis Using Generic Rate Equations*. *Metabolites*, 2:382–397, 2012.
- [2] ALONSO, A. P, D. L. VAL und Y. SHACHAR-HILL: *Central metabolic fluxes in the endosperm of developing maize seeds and their implications for metabolic engineering*. *Metabolic Engineering*, 13:96–107, 2011.
- [3] ALVES, R., F. ANTUNES und A. SALVADOR: *Tools for kinetic modeling of biochemical networks*. *Nature Biotechnology*, 24:667–672, 2006.
- [4] AMTHOR, J. S.: *From sunlight to phytomass: on the potential efficiency of converting solar radiation to phyto-energy*. *New Phytologist*, 188:939–959, 2010.
- [5] AREND, D., M. LANGE, J. CHEN, C. COLMSEE, S. FLEMMING, D. HECHT und U. SCHOLZ: *e!DAL - a framework to store, share and publish research data*. *BMC Bioinformatics*, 15:214–227, 2014.
- [6] ARISTOTELES: *Die Analytica priora (Erste Analytik) des Aristoteles ist die dritte Schrift des Organon und der erste Teil der Analytiken*. 350 BC.
- [7] BADER, G. D., M. P. CARY und C. SANDER: *Pathguide: a Pathway Resource List*. *Nucleic Acids Research*, 34:D504–D506, 2006.
- [8] BALDAN, P., N. COCCO, A. MARIN und M. SIMEONI: *Petri nets for modelling metabolic pathways: a survey*. *Natural Computing*, 9:955–989, 2010.
- [9] BASTIAN, M., S. HEYMANN und M. JACOMY: *Gephi: An Open Source Software for Exploring and Manipulating Networks*. In: *International AAAI Conference on Weblogs and Social Media*, Band 8 der Reihe *ICWSM '09*, Seiten 361–362. The AAAI Press, 2009.
- [10] BERESTOVSKY, N., W. ZHOU, D. NAGRATH und L. NAKHLEH: *Modeling Integrated Cellular Machinery Using Hybrid Petri-Boolean Networks*. *PLoS Computational Biology*, 9:e1003306, 2013.
- [11] BERG, J. M., J. L. TYMOCZKO, L. STRYER, B. HÄCKER und A. HELD: *Biochemie*. Spektrum Akademischer Verlag, 2003.
- [12] BERGMANN, F. T. und H. M. SAURO: *SBW - a Modular Framework for Systems Biology*. In: *Proceedings of the 38th Conference on Winter Simulation.*, WSC '06, Seiten 1637–1645. Winter Simulation Conference, 2006.
- [13] BERRY, J.A. und G.D. FARQUHAR: *The CO<sub>2</sub> concentrating function of C<sub>4</sub> photosynthesis: a biochemical model*. *Photosynthesis*, 77:119–131, 1978.

- [14] BIRCH, E. W., M. UDELL und M. W. COVERT: *Incorporation of flexible objectives and time-linked simulation with flux balance analysis*. Journal of Theoretical Biology, 345:12–21, 2014.
- [15] BISSWANGER, H.: *Enzymkinetik*. Wiley, 2000.
- [16] BRANDES, U., M. EIGLSPERGER, J. LERNER und C. PICH: *Graph Markup Language (GraphML)*. In: *Handbook of Graph Drawing and Visualization*, Seiten 517–541. CRC Press, 2013.
- [17] BRANDES, U. und D. WAGNER: *Visone-Analysis and Visualization of Social Networks*. In: *Graph Drawing Software*, Seiten 321–340. Springer-Verlag, 2004.
- [18] BÜCHEL, F., N. RODRIGUEZ, N. SWAINSTON, C. WRZODEK, T. CZAUDERNA, R. KELLER, F. MITTAG, M. SCHUBERT, M. GLONT, M. GOLEBIEWSKI et al.: *Path2Models: large-scale generation of computational models from biochemical pathway maps*. BMC Systems Biology, 7:116–134, 2013.
- [19] BURTON, A. C.: *The properties of the steady state compared to those of equilibrium as shown in characteristic biological behavior*. Journal of Cellular and Comparative Physiology, 13:327–349, 1939.
- [20] CALZONE, L., F. FAGES und S. SOLIMAN: *BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge*. Bioinformatics, 22:1805–1807, 2006.
- [21] CANTOR, G., R. DEDEKIND, E. ZERMELO und A. A. FRAENKEL: *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*. Verlag von Julius Springer, 1932.
- [22] CHAOUIYA, C., E. REMY und D. THIEFFRY: *Petri net modelling of biological regulatory networks*. Journal of Discrete Algorithms, 6:165–177, 2008.
- [23] CHEN, M., S. HARIHARAPUTRAN, R. HOFESTÄDT, B. KORMEIER und S. SPANGARDT: *Petri Net Models for the Semi-automatic Construction of Large Scale Biological Networks*. Natural Computing, 10:1077–1097, 2011.
- [24] CHOWDHURY, A., A. R. ZOMORRODI und C. D. MARANAS: *k-OptForce: Integrating Kinetics with Flux Balance Analysis for Strain Design*. PLoS Computational Biology, 10:e1003487, 2014.
- [25] COLLATZ, G.J., M. RIBAS-CARBO und J.A. BERRY: *Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C<sub>4</sub> Plants*. Australian Journal of Plant Physiology, 19:519–538, 1992.
- [26] COURTOT, M., N. JUTY, C. KNÜPFER, D. WALTEMATH, A. ZHUKOVA, A. DRÄGER, M. DUMONTIER, A. FINNEY, M. GOLEBIEWSKI, J. HASTINGS et al.: *Controlled vocabularies and semantics in systems biology*. Molecular Systems Biology, 7:543–554, 2011.
- [27] COVERT, M. W., C. H. SCHILLING und B. PALSSON: *Regulation of gene expression in flux balance models of metabolism*. Journal of Theoretical Biology, 213:73–88, 2001.

- 
- [28] COVERT, M. W., N. XIAO, T. J. CHEN und J. R. KARR: *Integrating Metabolic, Transcriptional Regulatory and Signal Transduction Models in Escherichia Coli*. *Bioinformatics*, 24:2044–2050, 2008.
- [29] CRICK, F. H. C.: *On Protein Synthesis*. Symposium of the Society of Experimental Biology, 12:138–163, 1958.
- [30] CRICK, F. H. C.: *Central Dogma of Molecular Biology*. *Nature*, 227:561–563, 1970.
- [31] CVIJOVIC, M., R. OLIVARES-HERNAACUTENDEZ, R. AGREN, N. DAHR, W. VONGSANGNAK, I. NOOKAEW, K. R. PATIL und J. NIELSEN: *BioMet Toolbox: genome-wide analysis of metabolism*. *Nucleic Acids Research*, 38:144–149, 2010.
- [32] CZAUDERNA, T., C. KLUKAS und F. SCHREIBER: *Editing, Validating, and Translating of SBGN Maps*. *Bioinformatics*, 26:2340–2341, 2010.
- [33] DANDEKAR, T., A. FIESELMANN, S. MAJEED und Z. AHMED: *Software applications toward quantitative metabolic flux analysis and modeling*. *Briefings in Bioinformatics*, 15:91–107, 2012.
- [34] DE JONG, H.: *Modeling and simulation of genetic regulatory systems: A literature review*. *Journal of Computational Biology*, 9:67–103, 2002.
- [35] DE OLIVEIRA DAL’MOLIN, C. G., L.-E. QUEK, R. W. PALFREYMAN, S. M. BRUMBLEY und L. K. NIELSEN: *C4GEM, a genome-scale metabolic model to study C4 plant metabolism*. *Plant Physiology*, 154:1871–1885, 2010.
- [36] DE VEAU, E. J. und J. E. BURRIS: *Photorespiratory Rates in Wheat and Maize as Determined by 18O-Labeling*. *Plant Physiology*, 90:500–511, 1989.
- [37] DEMIR, E., M. P. CARY, S. PALEY, K. FUKUDA, C. LEMER, I. VASTRIK, G. WU, P. D’EUSTACHIO, C. SCHAEFER, J. LUCIANO et al.: *The BioPAX community standard for pathway data sharing*. *Nature Biotechnology*, 28:935–942, 2010.
- [38] DHAR, P., T. C. MENG, S. SOMANI, L. YE, A. SAIRAM, M. CHITRE, H. ZHU und K. R. SAKHARKAR: *Cellware - a multi-algorithmic software for computational systems biology*. *Bioinformatics*, 20:1319–1321, 2004.
- [39] DIESTEL, R.: *Graph Theory*. Electronic Library of Mathematics. Springer-Verlag, 2000.
- [40] DINGLE, N. J., W. J. KNOTTENBELT und T. SUTO: *PIPE2: A Tool for the Performance Evaluation of Generalised Stochastic Petri Nets*. *ACM SIGMETRICS - Performance Evaluation Review*, 36:34–39, 2009.
- [41] DROSTE, P., K. NÖH und W. WIECHERT: *Omix: A Visualization Tool for Metabolic Networks with Highest Usability and Customizability in Focus*. *Chemie Ingenieur Technik*, 85:849–862, 2013.
- [42] DUNCAN, W. G. und A. L. HATFIELD: *A Method for Measuring the Daily Growth of Corn Kernels*. *Journal of Crop Science*, 4:550–551, 1964.
- [43] ECKLEDER, A. und T. FREYTAG: *WoPeD 2.0 goes BPEL 2.0*. In: *Proceedings of the 15th Workshop Algorithms and Tools for Petri Nets.*, Band 380 der Reihe *CEUR Workshop Proceedings*, Seiten 75–80. CEUR-WS.org, 2008.

- [44] EDWARDS, J. S., M. COVERT und B. O. PALSSON: *Metabolic modelling of microbes: the flux-balance approach*. Environmental Microbiology, 4:133–140, 2002.
- [45] EDWARDS, J. S. und B. O. PALSSON: *Robustness Analysis of the Escherichia coli Metabolic Network*. Biotechnology Progress, 16:927–939, 2000.
- [46] EICK, S. G. und G. J. WILLS: *High interaction graphics*. European Journal of Operational Research, 81:445–459, 1995.
- [47] FEIST, A. M. und B. O. PALSSON: *The biomass objective function*. Current Opinion in Microbiology, 13:344–349, 2010.
- [48] FISHER, D.: *Animation for Visualization: Opportunities and Drawbacks*. In: *Beautiful Visualization: Looking at Data through the Eyes of Experts*, Theory in Practice Series, Seiten 329–352. O’Reilly Media, 2010.
- [49] FRIEDRICHS, F. D.: *Referenznetze mit Anschriften in Scheme*. Diplomarbeit, Universität Hamburg, Fakultät für Mathematik, Informatik und Naturwissenschaften, 2007.
- [50] FUNAHASHI, A., Y. MATSUOKA, J. AKIYA, M. MOROHASHI, N. KIKUCHI und H. KITANO: *CellDesigner 3.5: A Versatile Modeling Tool for Biochemical Networks*. Proceedings of the IEEE, 96:1254–1265, 2008.
- [51] FURBANK, R. T.: *Evolution of the C<sub>4</sub> photosynthetic mechanism: are there really three C<sub>4</sub> acid decarboxylation types?* Journal of Experimental Botany, 62:3103–3108, 2011.
- [52] GARNY, A., D. P. NICKERSON, J. COOPER, R. WEBER DOS SANTOS, A. K. MILLER, S. MCKEEVER, P. M. F. NIELSEN und P. J. HUNTER: *CellML and associated tools and techniques*. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366:3017–3043, 2008.
- [53] GEIGENBERGER, P., M. STITT und A. R. FERNIE: *Metabolic control analysis and regulation of the conversion of sucrose to starch in growing potato tubers*. Plant, Cell & Environment, 27:655–673, 2004.
- [54] GEVORGYAN, A., M. E. BUSHHELL, C. AVIGNONE-ROSSA und A. M. KIERZEK: *SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks*. Bioinformatics, 27:433–434, 2011.
- [55] GILBERT, D. und M. HEINER: *From Petri Nets to Differential Equations - An Integrative Approach for Biochemical Network Analysis*. In: *Proceedings of the 27th International Conference on Applications and Theory of Petri Nets.*, Band 4024 der Reihe *Lecture Notes in Computer Science*, Seiten 181–200. Springer-Verlag, 2006.
- [56] GILBERT, D. R., H. FUSS, X. GU, R. J. ORTON, S. ROBINSON, V. VYSHEMIRSKY, M. J. KURTH, C. S. DOWNES und W. DUBITZKY: *Computational methodologies for modelling, analysis and simulation of signalling networks*. Briefings in Bioinformatics, 7:339–353, 2006.

- 
- [57] GONCALVES, E., J. BUCHER, A. RYLL, J. NIKLAS, K. MAUCH, S. KLAMT, M. ROCHA und J. SAEZ-RODRIGUEZ: *Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models*. *Molecular BioSystems*, 9:1576–1583, 2013.
- [58] GOOS, G.: *Vorlesungen über Informatik: Band 1: Grundlagen und funktionales Programmieren*. Springer-Lehrbuch. Springer-Verlag, 1997.
- [59] GOTTLÖB, F.: *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Verlag von Louis Nebert, 1879.
- [60] GRAFAHREND-BELAU, E.: *Rekonstruktion, Analyse und Visualisierung pflanzenbiologischer Stoffwechselprozesse*. Doktorarbeit, Martin-Luther-Universität Halle-Wittenberg, Naturwissenschaftliche Fakultät III, 2013.
- [61] GRAFAHREND-BELAU, E., A. JUNKER, A. ESCHENRÖDER, J. MÜLLER, F. SCHREIBER und B. H. JUNKER: *Multiscale Metabolic Modeling: Dynamic Flux Balance Analysis on a Whole-Plant Scale*. *Plant Physiology*, 163:637–647, 2013.
- [62] GRAFAHREND-BELAU, E., C. KLUKAS, B.H. JUNKER und F. SCHREIBER: *FBA-SimVis: interactive visualization of constraint-based metabolic models*. *Bioinformatics*, 25:2755–2757, 2009.
- [63] GRAFAHREND-BELAU, E., F. SCHREIBER, D. KOSCHÜTZKI und B. H. JUNKER: *Flux balance analysis of barley seeds: a computational approach to study systemic properties of central metabolism*. *Plant Physiology*, 149:585–598, 2009.
- [64] GRÄSSLER, J., D. KOSCHÜTZKI und F. SCHREIBER: *CentiLib: comprehensive analysis and exploration of network centralities*. *Bioinformatics*, 28:1178–1179, 2012.
- [65] GRIESSBACH, J.: *Erweiterung von VANTED zur Simulation und Visualisierung kinetischer Modelle*. Masterarbeit, Martin-Luther-Universität Halle-Wittenberg, Naturwissenschaftliche Fakultät III, 2014.
- [66] GULDBERG, C. M. und P. WAAGE: *Concerning Chemical Affinity*. *Erdmann's Journal für Praktische Chemie*, 127:69–114, 1879.
- [67] HARTMANN, A., H. ROHN, K. PUCKNAT und F. SCHREIBER: *Petri Nets in VANTED: Simulation of Barley Seed Metabolism*. In: *Proceedings of the 3rd International Workshop on Biological Processes & Petri Nets (BioPPN), satellite event of PETRI NETS 2012.*, Band 852 der Reihe *CEUR Workshop Proceedings*, Seiten 20–28. CEUR-WS.org, 2012.
- [68] HEINER, M., M. HERAJY, F. LIU, C. ROHR und M. SCHWARICK: *Snoopy - a unifying Petri net tool*. In: *Petri Nets*, Band 7347 der Reihe *Lecture Notes in Computer Science*, Seiten 398–407. Springer-Verlag, 2012.
- [69] HILLAH, L. M., E. KINDLER, F. KORDON, L. PETRUCCI und N. TREVES: *A primer on the Petri Net Markup Language and ISO/IEC 15909-2*. *Petri Net Newsletter*, 76:9–28, 2009.
- [70] HIMSOLT, M.: *GML: A portable Graph File Format*. Technischer Bericht, Universität Passau, 1996.
-

- [71] HOOPS, S., S. SAHLE, R. GAUGES, C. LEE, J. PAHLE, N. SIMUS, M. SINGHAL, L. XU, P. MENDES und U. KUMMER: *COPASI - a COmplex PATHway SIMulator*. *Bioinformatics*, 22:3067–3074, 2006.
- [72] HORST, R. J., G. DOEHLEMANN, R. WAHL, J. HOFMANN, A. SCHMIEDL, R. KAHMANN, J. KÄMPER und L. M. VOLL: *A model of Ustilago maydis leaf tumor metabolism*. *Plant Signaling and Behavior*, 5:1446–1449, 2010.
- [73] HÜBNER, K., S. SAHLE und U. KUMMER: *Applications and trends in systems biology in biochemistry*. *The FEBS Journal*, 278:2767–2857, 2011.
- [74] HUCKA, M., A. FINNEY, H. M. SAURO, H. BOLOURI, J. C. DOYLE, H. KITANO, A. P. ARKIN, B. J. BORNSTEIN, D. BRAY, A. CORNISH-BOWDEN et al.: *The Systems Biology Markup Language (SBML): A Medium For Representation And Exchange Of Biochemical Network Models*. *Bioinformatics*, 19:524–531, 2003.
- [75] JAMSHIDI, N. und B. O. PALSSON: *Mass Action Stoichiometric Simulation Models: Incorporating Kinetics and Regulation into Stoichiometric Models*. *Biophysical Journal*, 98:175–185, 2010.
- [76] JUNGNIKEL, D.: *Graphs, Networks and Algorithms*. Algorithms and Computation in Mathematics. Springer-Verlag, 2007.
- [77] JUNKER, B. H.: *Sucrose breakdown in the potato tuber*. Doktorarbeit, Universität Potsdam, Mathematisch-Naturwissenschaftliche Fakultät, 2004.
- [78] JUNKER, B. H., D. KOSCHÜTZKI und F. SCHREIBER: *Exploration of biological network centralities with CentiBiN*. *BMC Bioinformatics*, 7:219–225, 2006.
- [79] JUNKER, B. H. und F. SCHREIBER: *Analysis of Biological Networks*. Wiley Series in Bioinformatics. Wiley, 2011.
- [80] KARR, J. R., J. C. SANGHVI, D. N. MACKLIN, A. ARORA und M. W. COVERT: *WholeCellKB: model organism databases for comprehensive whole-cell models*. *Nucleic Acids Research*, 41:D787–D792, 2013.
- [81] KARR, J. R., J. C. SANGHVI, D. N. MACKLIN, M. V. GUTSCHOW, J. M. JACOBS, B. JR. BOLIVAL, N. ASSAD-GARCIA, J. I. GLASS und M. W. COVERT: *A Whole-Cell Computational Model Predicts Phenotype from Genotype*. *Cell*, 150:389–401, 2012.
- [82] KEITH, F. T. und B. SINÉAD: *History of the enzyme nomenclature system*. *Bioinformatics*, 16:34–40, 2000.
- [83] KELLER, R., A. DÖRR, A. TABIRA, A. FUNAHASHI, M. J. ZILLER, R. ADAMS, N. RODRIGUEZ, N. LE NOVÈRE, N. HIROI, H. PLANATSCHER et al.: *The systems biology simulation core algorithm*. *BMC Systems Biology*, 7:55–71, 2013.
- [84] KITANO, H.: *Computational systems biology*. *Nature*, 420:206–210, 2002.
- [85] KITANO, H.: *Systems biology: a brief overview*. *Science*, 295:1662–1664, 2002.
- [86] KLAMT, S., S. SCHUSTER und E. D. GILLES: *Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria*. *Biotechnology and Bioengineering*, 77:734–751, 2002.



- 
- [87] KLETZINSKY, V.: *Compendium der Biochemie*. Braumüller, 1858.
- [88] KLIPP, E., R. HERWIG, A. KOWALD, C. WIERLING und H. LEHRACH: *Systems Biology in Practice: Concepts, Implementation and Application*. Wiley, 2008.
- [89] KOCH, I., B. H. JUNKER und M. HEINER: *Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber*. *Bioinformatics*, 21:1219–1226, 2005.
- [90] KOLPAKOV, F., M. PUZANOV und A. KOSHUKOV: *BioUML: visual modeling, automated code generation and simulation of biological systems*. In: *Proceedings of the International Conference on Bioinformatics of Genome Regulation and Structure*, Band 3 der Reihe *BGRS '06*, Seiten 281–285, 2006.
- [91] KOSCHÜTZKI, D.: *Zentralitätsanalyse molekularbiologischer Netzwerke*. Doktorarbeit, Martin-Luther-Universität Halle-Wittenberg, Naturwissenschaftliche Fakultät III, 2011.
- [92] KOUNEV, S., S. SPINNER und P. MEIER: *QPME 2.0 - A Tool for Stochastic Modeling and Analysis Using Queueing Petri Nets*. In: *From Active Data Management to Event-Based Systems and More.*, Band 6462 der Reihe *Lecture Notes in Computer Science*, Seiten 293–311. Springer-Verlag, 2010.
- [93] KOZHENKOV, S., M. SEDOVA, Y. DUBININA, A. GUPTA, A. RAY, J. PONOMARENKO und M. BAITALUK: *Biological Networks - tools enabling the integration of multi-scale data for the host-pathogen studies*. *BMC Systems Biology*, 5:7–21, 2011.
- [94] KUMAR, S. P. und J. C. FEIDLER: *BioSPICE: A Computational Infrastructure for Integrative Biology*. *OMICS: A Journal of Integrative Biology*, 7:225–225, 2003.
- [95] KUMAR, V. S., M. S. DASIKA und C. D. MARANAS: *Optimization based automated curation of metabolic reconstructions*. *BMC Bioinformatics*, 8:212–228, 2007.
- [96] KUTZ, D.: *Editorerstellung für VANTED am Beispiel von SBML*. Diplomarbeit, Martin-Luther-Universität Halle-Wittenberg, Naturwissenschaftliche Fakultät III, 2012.
- [97] LAISK, A. und G. E. EDWARDS: *A mathematical model of C<sub>4</sub> photosynthesis: The mechanism of concentrating CO<sub>2</sub> in NADP-malic enzyme type species*. *Photosynthesis Research*, 66:199–224, 2000.
- [98] LAISK, A. und G. E. EDWARDS: *Leaf C<sub>4</sub> Photosynthesis in silico: The CO<sub>2</sub> Concentrating Mechanism*. In: *Understanding Complexity from Molecules to Ecosystems. Advances in Photosynthesis and Respiration*, Band 29, Seiten 323–348. Springer-Verlag, 2009.
- [99] LAKSHMANAN, M., G. KOH, B. K. S. CHUNG und D.-Y. LEE: *Software applications for flux balance analysis*. *Briefings in Bioinformatics*, 15:108–122, 2014.
- [100] LE NOVÈRE, N., M. HUCKA, H. MI, S. MOODIE, F. SCHREIBER, A. SOROKIN, E. DEMIR, K. WEGNER, M. I. ALADJEM, S. M. WIMALARATNE et al.: *The Systems Biology Graphical Notation*. *Nature Biotechnology*, 27:735–741, 2009.
-

- [101] LEE, J. M., E. P. GIANCHANDANI, J. A. EDDY und J. A. PAPIN: *Dynamic Analysis of Integrated Signaling, Metabolic, and Regulatory Networks*. PLoS Computational Biology, 4:e1000086, 2008.
- [102] LEE, Y., F. CHEN, L. GALLEG0-GIRALDO, R. A. DIXON und E. O. VOIT: *Integrative Analysis of Transgenic Alfalfa (Medicago sativa L.) Suggests New Metabolic Control Mechanisms for Monolignol Biosynthesis*. PLoS Computational Biology, 7:e002047, 2011.
- [103] LIAO, Y-C., M.-H. TSAI, F.-C. CHEN und C. A. HSIUNG: *GEMSiRV: a software platform for GEnome-scale metabolic model simulation, reconstruction and visualization*. Bioinformatics, 28:1752–1758, 2012.
- [104] LLANERAS, F. und J. PICO: *Stoichiometric modelling of cell metabolism*. Journal of Bioscience and Bioengineering, 105:1–11, 2008.
- [105] LUNN, J. E. und M. D. HATCH: *Primary partitioning and storage of photosynthate in sucrose and starch in leaves of C<sub>4</sub> plants*. Planta, 197:385–391, 1995.
- [106] MACHADO, D., R. COSTA, M. ROCHA, E. C. FERREIRA, B. TIDOR und I. ROCHA: *Modeling formalisms in Systems Biology*. AMB Express, 1:45–58, 2011.
- [107] MACHADO, D., R. S. COSTA, M. ROCHA, I. ROCHA, B. TIDOR und E. C. FERREIRA: *Model Transformation of Metabolic Networks using a Petri Net Based Framework*. In: DONATELLI, S., J. KLEIJN, R. J. MACHADO und J. M. FERNANDES (Herausgeber): *ACSD/Petri Nets Workshops*, Band 827 der Reihe *CEUR Workshop Proceedings*, Seiten 103–117. CEUR-WS.org, 2010.
- [108] MAHADEVAN, R. und C.H. SCHILLING: *The effects of alternate optimal solutions in constraint-based genome-scale metabolic models*. Metabolic Engineering, 5:264–276, 2003.
- [109] MAJERAN, W., G. FRISO, L. PONNALA, B. CONNOLLY, M. HUANG, E. REIDEL, C. ZHANG, Y. ASAKURA, N. H. BHUIYAN, Q. SUN et al.: *Structural and Metabolic Transitions of C<sub>4</sub> Leaf Development and Differentiation Defined by Microscopy and Quantitative Proteomics in Maize*. Plant Cell, 22:3509–3542, 2010.
- [110] MARSAN, M. A.: *Stochastic Petri Nets: An Elementary Introduction*. In: *Advances in Petri Nets*, Band 424 der Reihe *Lecture Notes in Computer Science*, Seiten 1–29. Springer-Verlag, 1990.
- [111] MARTIN, A. R. und M. O. WARD: *High Dimensional Brushing for Interactive Exploration of Multivariate Data*. In: *Proceedings of the 6th conference on Visualization '95, VIS '95*, Seiten 271–278. IEEE Computer Society, 1995.
- [112] MOODIE, S., N. LE NOVÈRE, E. DEMIR, H. MI und F. SCHREIBER: *Systems Biology Graphical Notation: Process Description language Level 1*. Available from Nature Preceedings <http://dx.doi.org/10.1038/npre.2011.3721.4>, 2011.
- [113] MURATA, T.: *Petri nets: Properties, Analysis and Applications*. Proceedings of the IEEE, 77:541–580, 1989.

- 
- [114] MYERS, C. J., N. BARKER, K. JONES, H. KUWAHARA, C. MADSEN und N.-P. D. NGUYEN: *iBioSim: a tool for the analysis and design of genetic circuits*. *Bioinformatics*, 25:2848–2849, 2009.
- [115] ORTH, J. D., I. THIELE und B. O. PALSSON: *What is flux balance analysis?* *Nature Biotechnology*, 28:245–248, 2010.
- [116] PALSSON, B. O.: *Systems Biology: Properties of Reconstructed Networks*. Cambridge University Press, 2006.
- [117] PALSSON, B. O.: *Systems Biology: Simulation of Dynamic Network States*. Cambridge University Press, 2011.
- [118] PFAU, T., N. CHRISTIAN und O. EBENHÖH: *Systems approaches to modelling pathways and networks*. *Briefings in Functional Genomics*, 10:266–279, 2011.
- [119] PICK, T.R., A. BRÄUTIGAM, U. SCHLÜTER, A.K. DENTON, C. COLMSEE, U. SCHOLZ, H. FAHNENSTICH, R. PIERUSCHKA, U. RASCHER et al.: *Systems Analysis of a Maize Leaf Developmental Gradient Redefines the Current C4 Model and Provides Candidates for Regulation*. *The Plant Cell*, 23:4208–4220, 2011.
- [120] PILALIS, E., A. CHATZIOANNOU, B. THOMASSET und F. KOLISIS: *An in-silico compartmentalized metabolic model of Brassica napus enables the systemic study of regulatory aspects of plant central metabolism*. *Biotechnology and Bioengineering*, 108:1673–1682, 2011.
- [121] PUCKNAT, K.: *Erweiterung der Software VANTED zum Editieren und Simulieren biologischer Petri-Netze*. Masterarbeit, Hochschule Ostwestfalen-Lippe, Fachbereich Umweltingenieurwesen und Angewandte Informatik, 2012.
- [122] QU, Z., A. GARFINKEL, J. N. WEISS und M. NIVALA: *Multi-scale modeling in biology: How to bridge the gaps between scales?* *Progress in Biophysics and Molecular Biology*, 107:21–31, 2011.
- [123] RAMSEY, S., D. ORRELL und H. BOLOURI: *Dizzy: stochastic simulation of large-scale genetic regulatory networks*. *Journal of Bioinformatics and Computational Biology*, 3:415–436, 2005.
- [124] REED, J. L., T. D. VO, C. H. SCHILLING und B. O. PALSSON: *An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)*. *Genome Biology*, 4:R54.1–R54.12, 2003.
- [125] RESAT, H., L. PETZOLD und M. F. PETTIGREW: *Kinetic Modeling of Biological Systems*. In: *Computational Systems Biology*, Band 541 der Reihe *Methods in Molecular Biology*, Seiten 311–335. Humana Press, 2009.
- [126] ROCHA, I., P. MAIA, P. EVANGELISTA, P. VILACA, S. SOARES, J. PINTO, J. NIELSEN, K. PATIL, E. FERREIRA und M. ROCHA: *OptFlux: an open-source software platform for in silico metabolic engineering*. *BMC Systems Biology*, 4:45, 2010.

- [127] RODRIGUEZ, R. J., J. JULVEZ und J. MERSEGUER: *Peabrain: A PIPE Extension for Performance Estimation and Resource Optimisation*. In: *Proceedings of the 12th International Conference on Application of Concurrency to System Design.*, Seiten 142–147. IEEE Computer Society, 2012.
- [128] ROHN, H., A. JUNKER, A. HARTMANN, E. GRAFAHREND-BELAU, H. TREUTLER, M. KLAPPERSTÜCK, T. CZAUDERNA, C. KLUKAS und F. SCHREIBER: *VANTED v2: a framework for systems biology applications*. BMC Systems Biology, 6:139–152, 2012.
- [129] ROHR, C.: *Simulative Model Checking of Steady-State and Time-Unbounded Temporal Operators*. In: *Proceedings of the 3rd International Workshop on Biological Processes & Petri Nets (BioPPN), satellite event of PETRI NETS 2012.*, Band 852 der Reihe CEUR Workshop Proceedings, Seiten 62–75. CEUR-WS.org, 2012.
- [130] ROHWER, J. M.: *Kinetic modelling of plant metabolic pathways*. Journal of Experimental Botany, 63:2275–2292, 2012.
- [131] ROHWER, J. M. und F. C. BOTHA: *Analysis of sucrose accumulation in the sugar cane culm on the basis of in vitro kinetic data*. Biochemical Journal, 358:437–445, 2001.
- [132] SAGE, R. F.: *The evolution of C<sub>4</sub> photosynthesis*. New Phytologist, 161:341–370, 2004.
- [133] SAHA, R., P. F. SUTHERS und C. D. MARANAS: *Zea mays iRS1563: A Comprehensive Genome-Scale Metabolic Reconstruction of Maize Metabolism*. PLoS ONE, 6:e21784, 2011.
- [134] SAIER JR., M. H., M. R. YEN, K. NOTO, D. G. TAMANG und C. ELKAN: *The Transporter Classification Database: recent advances*. Nucleic Acids Research, 37:D274–D278, 2009.
- [135] SBML EDITORS: *The SBML Software Guide*. Available on the World Wide Web at [http://sbml.org/SBML\\_Software\\_Guide/](http://sbml.org/SBML_Software_Guide/), 2014.
- [136] SBML EDITORS: *The SBML Specifications*. Available on the World Wide Web at <http://sbml.org/Documents/Specifications/>, 2014.
- [137] SCHALLAU, K. und B. H. JUNKER: *Simulating plant metabolic pathways with enzyme-kinetic models*. Plant Physiology, 152:1763–1771, 2010.
- [138] SCHILLING, C. H., D. LETSCHER und B. O. PALSSON: *Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective*. Journal of Theoretical Biology, 203:229–248, 2000.
- [139] SCHREIBER, F.: *Visualisierung biochemischer Reaktionsnetze*. Doktorarbeit, Universität Passau, Fakultät für Mathematik und Informatik, 2001.
- [140] SCHREIBER, F., C. COLMSEE, T. CZAUDERNA, E. GRAFAHREND-BELAU, A. HARTMANN, A. JUNKER, B. H. JUNKER, M. KLAPPERSTÜCK, U. SCHOLZ und S. WEISE: *MetaCrop 2.0: managing and exploring information about crop plant metabolism*. Nucleic Acids Research, 40:D1173–D1177, 2012.

- 
- [141] SCHUETZ, R., L. KUEPFER und U. SAUER: *Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli*. *Molecular Systems Biology*, 3:119–134, 2007.
- [142] SCHUSTER, S. und C. HILGETAG: *On Elementary Flux Modes in Biochemical Reaction Systems at Steady State*. *Journal of Biological Systems*, 2:165–182, 1994.
- [143] SCHUSTER, S. und B. H. JUNKER: *Topological analysis of metabolic and regulatory networks*. In: *Modeling in Systems Biology: The Petri Net Approach.*, Band 16 der Reihe *Computational Biology*, Seiten 209–224. Springer-Verlag, 2011.
- [144] SCHWARZ, R., C. LIANG, C. KALETA, M. KÜHNEL, E. HOFFMANN, S. KUZNETSOV, M. HECKER, G. GRIFFITHS, S. SCHUSTER und T. DANDEKAR: *Integrated network reconstruction, visualization and analysis using YANAsquare*. *BMC Bioinformatics*, 8:313–323, 2007.
- [145] SHANNON, P., A. MARKIEL, O. OZIER, N. S. BALIGA, J. T. WANG, D. RAMAGE, N. AMIN, B. SCHWIKOWSKI und T. IDEKER: *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. *Genome Research*, 13:2498–2504, 2003.
- [146] SHLOMI, T., Y. EISENBERG, R. SHARAN und E. RUPPIN: *A genome-scale computational study of the interplay between transcriptional regulation and metabolism*. *Molecular Systems Biology*, 3:101, 2007.
- [147] SHNEIDERMAN, B.: *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In: *Proceedings of the 1996 IEEE Symposium on Visual Languages*, Visual Languages '96, Seiten 336–344. IEEE Computer Society, 1996.
- [148] SIMONS, M. N., R. SAHA, N. AMIOUR, A. KUMAR, L. GUILLARD, G. S. CLÉMENT, M. MIQUEL, Z. LI, G. MOUILLE, P. J. LEA et al.: *Assessing the Metabolic Impact of Nitrogen Availability using a Compartmentalized Maize Leaf Genome-Scale Model*. *Plant Physiology*, pp.114.245787, 2014.
- [149] SMALLBONE, K., E. SIMEONIDIS, D. S. BROOMHEAD und D. B. KELL: *Something from nothing – bridging the gap between constraint-based and kinetic modelling*. *FEBS Journal*, 274:5576–5585, 2007.
- [150] SMALLBONE, K., E. SIMEONIDIS, N. SWAINSTON und P. MENDES: *Towards a genome-scale kinetic model of cellular metabolism*. *BMC Systems Biology*, 4:6–18, 2010.
- [151] STARKE, P. H.: *Analyse von Petri-Netz Modellen*. Leitfäden und Monographien der Informatik. Teubner Verlag, 1990.
- [152] STEPHANOPOULOS, G.: *Metabolic Fluxes and Metabolic Engineering*. *Metabolic Engineering*, 1:1–11, 1999.
- [153] STEUER, R und B. H. JUNKER: *Computational Models of Metabolism: Stability and Regulation in Metabolic Networks*. In: *Advances in Chemical Physics*, Band 142, Seiten 105–251. John Wiley and Sons, Inc., 2008.

- [154] STREHMEL, K., R. WEINER und H. PODHAISKY: *Numerik gewöhnlicher Differentialgleichungen: Nichtsteife, steife und differential-algebraische Gleichungen*. Vieweg+Teubner Verlag, 2012.
- [155] TENAZINHA, N. und S. VINGA: *A Survey on Methods for Modeling and Analyzing Integrated Biological Networks*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 8:943–958, 2011.
- [156] THOMAS, J. J. und K. A. COOK: *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005.
- [157] UNIVERSITÄT DÜSSELDORF: *Querschnitt durch das Blatt einer Maispflanze*. Available on the World Wide Web at [http://www.uni-duesseldorf.de/MathNat/Biologie/Didaktik/Fotosynthese\\_neu/dateien/dunkel/bilder/maisquer.jpg](http://www.uni-duesseldorf.de/MathNat/Biologie/Didaktik/Fotosynthese_neu/dateien/dunkel/bilder/maisquer.jpg), 2014.
- [158] VAN IERSEL, M. P., A. C. VILLÉGER, T. CZAUDERNA, S. E. BOYD, F. T. BERGMANN, A. LUNA, E. DEMIR, A. SOROKIN, U. DOGRUSOZ, Y. MATSUOKA et al.: *Software support for SBGN maps: SBGN-ML and LibSBGN*. Bioinformatics, 28:2016–2021, 2012.
- [159] VOIT, E. O.: *Biochemical Systems Theory: A Review*. ISRN Biomathematics, Seiten 1–53, 2013.
- [160] VON CAEMMERER, S.: *Biochemical Models of Leaf Photosynthesis*. Techniques in Plant Sciences. CSIRO Publishing, 2000.
- [161] VON LANDESBERGER, T., A. KUIJPER, T. SCHRECK, J. KOHLHAMMER, J. VAN WIJK, J.-D. FEKETE und D. W. FELLNER: *Visual Analysis of Large Graphs: State-of-the-Art and Future Research Challenges*. Computer Graphics Forum, 30:1719–1749, 2011.
- [162] WANG, C., L. GUO, Y. LI und Z. WANG: *Systematic Comparison of C<sub>3</sub> and C<sub>4</sub> Plants Based on Metabolic Network Analysis*. BMC Systems Biology, 6:S9, 2012.
- [163] WANG, Y., S. P. LONG und X.-G. ZHU: *Elements Required for an efficient NADP-ME type C<sub>4</sub> photosynthesis*. Plant Physiology, 164:2231–2246, 2014.
- [164] WARD, M., G. GRINSTEIN und D. KEIM: *Interactive Data Visualization: Foundations, Techniques, and Applications*. A. K. Peters, Ltd., 2010.
- [165] WEBER, A. P. M. und S. VON CAEMMERER: *Plastid transport and metabolism of C<sub>3</sub> and C<sub>4</sub> plants comparative analysis and possible biotechnological exploitation*. Current Opinion in Plant Biology, 13:256–264, 2010.
- [166] WHITE, P. J. und L. A. JOHNSON: *Corn: chemistry and technology*. American Association of Cereal Chemists, 2 Auflage, 2003.
- [167] WIECHERT, W.: *Modeling and simulation: tools for metabolic engineering*. Journal of Biotechnology, 94:37–63, 2002.
- [168] WRIGHT, J. und A. WAGNER: *The Systems Biology Research Tool: evolvable open-source software*. BMC Systems Biology, 2:55–61, 2008.

- [169] XU, C., L. LIU, Z. ZHANG, D. JIN, J. QIU und M. CHEN: *Genome-scale metabolic model in guiding metabolic engineering of microbial improvement*. Applied Microbiology and Biotechnology, 97:519–539, 2013.
- [170] ZHOU, T., J. REN, M. MEDO und Y. C. ZHANG: *Bipartite network projection and personal recommendation*. Physical Review E (Statistical, Nonlinear, and Soft Matter Physics), 76:46115–46122, 2007.





# Publikationen

Bereiche der Dissertation, die in Publikationen einfließen, sind mit einem Doppelkreuz (#) hervorgehoben.

1. HARTMANN, A. und F. SCHREIBER: *Integrative Analysis of Metabolic Models - from Structure to Dynamics*. Revision submitted to Journal: *Frontiers in Bioengineering and Biotechnology*, Section: *Systems Biology*, Topic: *Current challenges in modeling cellular metabolism*. #
2. KUNZE, G., C. GAILLARDIN, M. CZERNICKA, P. DURRENS, T. MARTIN, E. BÖER, T. GABALDPN, J. A. CRUZ, E. TALLA, C. MARCK, A. GOFFEAU, V. BARBE, P. BARET, K. BARONIAN, S. BEIER, C. BLEYKASTEN, R. BODE, S. CASAREGOLA, L. DESPONS, C. FAIRHEAD, M. GIERSBERG, P. P. GIERSKI, U. HÄHNEL, A. HARTMANN, D. JANKOWSKA, C. JUBIN, P. JUNG, I. LAFONTAINE, V. LEH-LOUIS, M. LEMAIRE, M. MARCET-HOUBEN, M. MASCHER, G. MOREL, G.-F. RICHARD, J. RIECHEN, C. SACERDOT, A. SARKAR, G. SAVEL, J. SCHACHERER, D. J. SHERMAN, N. STEIN, M.-L. STRAUB, A. THIERRY, A. TRAUTWEIN-SCHULT, B. VACHERIE, E. WESTHOF, S. WORCH, B. DUJON, J.-L. SOUCIET, P. WINCKER, U. SCHOLZ und C. NEUVEGLISE: *The complete genome of Blastobotrys (Arxula) adenivorans LS3 - a yeast of biotechnological interest*. *Biotechnology for Biofuels*, 7:66–81, 2014.
3. LOTZ, K.\* , A. HARTMANN\*, E. GRAFAHREND-BELAU, F. SCHREIBER und B. H. JUNKER: *Elementary flux modes, flux balance analysis, and their application to plant metabolism*. In: *Sriram G (Ed.): Plant Metabolism: Methods and Protocols. (Series: Methods in Molecular Biology, Vol. 1083)*. Humana Press, Totowa, New Jersey, 2014, \*equally contributing. #
4. COLMSEE, C., T. CZAUDERNA, E. GRAFAHREND-BELAU, A. HARTMANN, M. LANGE, M. MASCHER, S. WEISE, U. SCHOLZ und F. SCHREIBER: *OPTIMAS-DW, MetaCrop and VANTED: a case study for data integration, curation and visualisation in life sciences*. In: *Proceedings of Ontologies and Data in Life Sciences*, Band P-220 der Reihe *Lecture Notes in Informatics*, Seiten 1834–1840, 2013. #
5. COLMSEE, C., M. MASCHER, T. CZAUDERNA, A. HARTMANN, U. SCHLÜTER, N. ZELLERHOFF , J. SCHMITZ, A. BRÄUTIGAM, T. R. PICK, P. ALTER, M. GAHRTZ, S. WITT, A. R. FERNIE, F. BÖRNKE, H. FAHNENSTICH, M. BUCHER, T. DRESSELHAUS, A. P. WEBER, F. SCHREIBER, U. SCHOLZ und U. SONNEWALD: *OPTIMAS-DW: A comprehensive transcriptomics, metabolomics, ionomics, proteomics and phenomics data resource for maize*. *BMC Plant Biology* 12:245–255, 2012. #

6. HARTMANN, A., H. ROHN, K. PUCKNAT und F. SCHREIBER: *Petri nets in VANTED: Simulation of barley seed metabolism*. In: *Proceedings of the 3rd International Workshop on Biological Processes & Petri Nets (BioPPN), satellite event of PETRI NETS 2012.*, Band 852 der Reihe *CEUR Workshop Proceedings*, Seiten 20–28, 2012. <#>
7. JUNKER, A., H. ROHN, T. CZAUDERNA, C. KLUKAS, A. HARTMANN und F. SCHREIBER: *Creating interactive, web-based and data-enriched maps using the Systems Biology Graphical Notation*. *Nature Protocols* 7:579–593, 2012. <#>
8. ROHN H, A. HARTMANN, A. JUNKER, B. H. JUNKER und F. SCHREIBER: *FluxMap: A VANTED add-on for the visual exploration of flux distributions in biological networks*. *BMC Systems Biology*, 6:33–42, 2012. <#>
9. ROHN, H., A. JUNKER, A. HARTMANN, E. GRAFAHREND-BELAU, H. TREUTLER, M. KLAPPERSTÜCK, T. CZAUDERNA, C. KLUKAS und F. SCHREIBER: *VANTED v2: a framework for systems biology applications*. *BMC Systems Biology*, 6:139–152, 2012. <#>
10. SCHREIBER, F., C. COLMSEE, T. CZAUDERNA, E. GRAFAHREND-BELAU, A. HARTMANN, A. JUNKER, B. H. JUNKER, M. KLAPPERSTÜCK, U. SCHOLZ und S. WEISE: *MetaCrop 2.0: managing and exploring information about crop plant metabolism*. *Nucleic Acids Research*, 40:D1173–D1177, 2012. <#>
11. HARTMANN, A., T. CZAUDERNA, R. HOFFMANN, N. STEIN und F. SCHREIBER: *HTPheno: An image analysis pipeline for high- throughput plant phenotyping*. *BMC Bioinformatics*, 12: 1–9, 2011.
12. JUNKER, A., A. HARTMANN, F. SCHREIBER und H. BÄUMLEIN: *An engineer's view on regulation of seed development*. *Trends in Plant Science*, 15:303–307, 2010.

# Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Datum

Unterschrift des Antragstellers



# Lebenslauf

## Anja Hartmann

Adresse: Weberstraße 42  
06484 Quedlinburg

E-Mail: anja.hartmann0406@gmail.com

Geboren am: 04.06.1983

Ort: Quedlinburg

Familienstand: ledig

Nationalität: deutsch

## Berufliche Tätigkeit

seit 08/2009 wissenschaftliche Mitarbeiterin und Doktorandin, Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) Gatersleben

*Projekt:* OPTIMAS–Systembiologische Modellierung  
der Ertragsbildung bei Mais

*Projekt:* SEA2–Systembiologische Steigerung  
des Ölertrages von Raps

## Ausbildung und Studium

10/2003–07/2009 Studium der Bioinformatik, Martin-Luther-Universität Halle-Wittenberg, Abschluss mit Diplom, Abschlussnote: 1,3

*Diplomarbeit:* Automatisierte Datenaufnahme und Bildverarbeitung zur Hochdurchsatz-Phänotypisierung von Gerste

bis 06/2003 Erlangung der Hochschulreife, Dorothea-Erxleben-Gymnasium Quedlinburg, Abschlussnote: 2,2

Datum

Unterschrift des Antragstellers