



Institut für Biotechnologie

***Erstellung und Validierung mechanistischer
Modelle für den mikrobiellen Stoffwechsel
zur Auswertung von Substrat-Puls-
Experimenten***

Daniela Degenring

***Erstellung und Validierung mechanistischer
Modelle für den mikrobiellen Stoffwechsel
zur Auswertung von Substrat-Puls-
Experimenten***

Daniela Degenring

Berichte des Forschungszentrums Jülich ; 4147

ISSN 0944-2952

Institut für Biotechnologie Jül-4147

D 28 (Diss., Rostock, Univ., 2004)

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek

D-52425 Jülich · Bundesrepublik Deutschland

☎ 02461 61-5220 · Telefax: 02461 61-6103 · e-mail: zb-publikation@fz-juelich.de

Danksagung

Die vorliegende Arbeit wurde in der Zeit von Oktober 2000 bis Oktober 2003 am Lehrstuhl für Technische Chemie der Universität Rostock in Kooperation mit der Fermentationsgruppe am Institut für Biotechnologie II des Forschungszentrums Jülich angefertigt.

Mein Dank gilt:

- Herrn Prof. Kragl, Lehrstuhl für Technische Chemie der Universität Rostock, für viele konstruktive Beiträge und die freundliche Bereitschaft, das Hauptreferat zu übernehmen. Sein engagierter Einsatz für die Arbeit und sein steter Optimismus wirkten motivierend und trugen zum Gelingen der Arbeit wesentlich bei.
- Herrn Prof. Wolkenhauer, Lehrstuhl für Bioinformatik und Systembiologie an der Universität Rostock, für die interessanten fachlichen Diskussionen und für seine freundliche Bereitschaft, ein Korreferat zu übernehmen.
- Herrn Prof. Wandrey, Leiter des Instituts für Biotechnologie II des Forschungszentrums Jülich und Inhaber des Lehrstuhls für Technische Chemie der Universität Bonn, für die Gelegenheit, an seinem Institut arbeiten zu dürfen, die Bereitstellung der Arbeitsmittel und seine freundliche Bereitschaft, ein Korreferat zu übernehmen.
- Herrn Dr. Takors, der in seiner Eigenschaft als Gruppenleiter der Fermentationsgruppe am Institut für Biotechnologie II des Forschungszentrums Jülich die Arbeit betreute und die dreijährige Anstellung im DFG-Projekt TA241/3-1 ermöglichte.
- Frau Prof. Uhrmacher, Leiterin der Arbeitsgruppe Modellierung und Simulation am Institut für Informatik der Universität Rostock, für ihre große Hilfs- und Diskussionsbereitschaft.
- Herrn Prof. Dikta, Professor für Mathematik und Angewandte Mathematik, Fachhochschule Aachen Abteilung Jülich, für das Interesse an der Arbeit und die fruchtbaren fachlichen Diskussionen.
- Herrn Prof. Wiechert, Lehrstuhl für Simulationstechnik der Universität Siegen und Herrn Prof. Freisleben, Arbeitsgruppe für verteilte Systeme des Fachbereichs Mathematik und Informatik der Universität Marburg, für die Kooperation innerhalb des DFG-Projektes.
- Dr. Jochen Hurlbaus, Dr. Arne Bucholz, Dipl.-Chem. Marco Oldiges, Dipl.-Ing. Marc-Daniel Haunschild für die experimentellen Arbeiten und softwaretechnischen Entwicklungen, ohne die die Arbeit in dieser Form nicht zustande gekommen wäre.
- Frau Dipl.-Math. Cornelia Froemel für die engagierte Bearbeitung mit dieser Arbeit zusammenhängender Fragestellungen im Rahmen ihrer Diplomarbeit.
- Den Mitarbeitern am Institut für Biotechnologie II für die freundliche Aufnahme am Institut.
- Den Mitarbeitern des Lehrstuhls Technische Chemie der Universität Rostock für die gute Arbeitsatmosphäre und die positive Unterstützung.
- Allen Mitarbeitern und Fritz der Arbeitsgruppe Modellierung und Simulation der Universität Rostock für die freundliche, angenehme Arbeitsatmosphäre.
- Allen bisher nicht genannten Personen, die zum Gelingen der Arbeit beigetragen haben.

Zusammenfassung

Ziel der vorliegenden Arbeit war es, valide Differentialgleichungsmodelle für die Beschreibung mikrobieller Stoffwechselvorgänge zu entwickeln. Grundlage der Modelle sind Datensätze aus Substrat-Puls-Experimenten, in denen verschiedene *Escherichia coli*-Stämme durch Glucose-Pulse angeregt und die Veränderungen der intrazellulären Metabolitkonzentrationen zeitlich hoch aufgelöst erfaßt wurden. Die gemessenen Datenreihen waren im Rahmen dieser Arbeit mit computergenerierten Modellen zu beschreiben sowie generelle Schwierigkeiten, die im Verlauf des Auswertungsprozesses auftreten können, zu identifizieren und methodisch anzugehen.

Dazu wurde der gesamte Modellierungs- und Simulationszyklus iterativ beschritten. Diesem Ansatz der explorativen Modellierung folgend, wurde eine große Anzahl an Differentialgleichungssystemen zur Auswertung der Meßdaten erstellt. Die Datensätze und Modelle wurden umfangreich validiert, indem bewährte Methoden aus dem Bereich der Modellierung & Simulation übernommen und angewendet wurden. Dies umfaßte eine Fehlerrechnung, die Berechnung von Bestimmungsgrenzen zur Prüfung der Datenqualität sowie Plausibilitätstests und den Einsatz einer induktiven, statistischen Methode zur Untersuchung der Konsistenz der Datensätze. Mehrere Arten für die Analyse der Parametersensitivitätsfunktionen wurden zur Reduktion der Parameterzahl sowie zur anschließenden Prüfung der Struktur und der praktischen Identifizierbarkeit der Modelle angewendet. Des weiteren wurden verschiedene Kriterien zur Modellselektion bzw. Diskriminierung auf Basis der Anpassungsgüte getestet. Die erfolgreichen Tests aller Methoden an zahlreichen Beispielsystemen und den gegebenen Datensätzen dokumentieren die Einsatzfähigkeit der Verfahren.

Die Erkenntnisse, welche Methode wann und wie angewandt werden kann, flossen in ein generalisiertes Konzept für einen Modellierungs- und Validierungsprozeß ein. Damit wurde ein erster Schritt hin zu einer Standardisierung des Auswertungs- und Beurteilungsprozesses von Substrat-Puls-Experimenten im Kontext der Systembiologie getan.

Inhaltsverzeichnis

1	Einleitung und Zielstellung	1
1.1	Einleitung	1
1.2	Zielstellung	2
1.3	Gliederung der Arbeit	3
2	Theoretische Grundlagen	5
2.1	Biologische Grundlagen	5
2.1.1	Beschreibung einer prokaryotischen Zelle	5
2.1.2	Der Mikroorganismus <i>Escherichia coli</i>	7
2.1.3	Metabolic Engineering	8
2.2	Modellierung der Zelle	8
2.2.1	Allgemeines	8
2.2.2	Modellierungsansätze zur Beschreibung metabolischer Netzwerke	10
2.2.3	Mechanistische Modelle zur Untersuchung der Stoffwechseldynamiken	14
2.3	Verifizieren, Validieren und Testen von Modellen	19
2.3.1	Allgemeines	19
2.3.2	Modellierungs- und Validierungsprozeß	20
3	Konzept für den Modellierungs- und Validierungsprozeß	25
3.1	Einleitung	25
3.2	Substrat-Puls-Experimente	25
3.2.1	Experiment mit einem <i>Escherichia coli</i> Wildstamm	27
3.2.2	Experimente mit Phenylalanin-Produktionsstämmen	27
3.2.3	Datensätze	29
3.3	Konzeptuelles Modell	30
3.3.1	Validierung der Daten	30
3.3.2	Hypothesen und vereinfachende Annahmen	32
3.4	Computermode ll	35
3.4.1	Erstellung und Verifikation des Computermode ll	35
3.4.2	Operationale Modellvalidierung	39
4	Validierung der Daten	43
4.1	Allgemeine Bemerkungen	43

4.2	Fehlerrechnung	43
4.2.1	Allgemeines	43
4.2.2	Fehlerrechnung für die intrazellulären Metabolitkonzentrationen	45
4.3	Nachweis-, Erfassungs- und Bestimmungsgrenzen	48
4.4	Standardisierung der Datenerfassung – Software PulseSTAT	50
4.5	Glättung der Daten	51
4.5.1	FFT-Filterung	52
4.5.2	Polynomische und Splineapproximationen	53
4.6	Plausibilitätstests	54
4.7	Zeitreihenanalyse	56
4.7.1	Einführung	56
4.7.2	Implementierung der Zeitreihenanalyse	58
4.7.3	Test des Verfahrens anhand von <i>in-silico</i> Datensätzen	60
4.7.4	Anwendung auf Datensätze aus Substrat-Puls-Experimenten	65
4.7.5	Erweiterungen der Methode	67
5	Erstellung und Validierung des konzeptuellen Modells	69
5.1	Einführung	69
5.2	Modelle zur Beschreibung der Aromatenbiosynthese in <i>E.coli</i> -Mutanten	69
5.2.1	Allgemeine Hypothesen und vereinfachende Annahmen	69
5.2.2	Modelle mit einem vorgegebenen Flußprofil als Randbedingung	71
5.2.3	Modelle mit der Vorgabe von Konzentrationsverläufen als Randbedingung	75
5.2.4	Zusammenfassung der Modellfamilie	77
5.3	Modelle für die Glykolyse und den Pentose-Phosphat-Weg von <i>E.coli</i> K12	77
5.3.1	Vorstellung eines Modells	77
5.3.2	Zusammenfassung der Ergebnisse aller Modelle für das Experiment WT1	84
6	Operationale Modellvalidierung	87
6.1	Einleitung	87
6.2	Sensitivitätsanalyse	87
6.2.1	Motivation	87
6.2.2	Lokale Sensitivitätsfunktionen	88
6.2.3	Implementierung der Sensitivitätsberechnung in MMT	90
6.2.4	Eliminierung der insignifikanten Parameter	91
6.2.5	Zusammenfassen korrelierender Parameter	99

6.2.6	Parameterschätzfehler	102
6.2.7	Zusammenfassung	104
6.3	Bootstrap-Datensätze für den Test der Anpassungsgüte	104
6.3.1	Einleitung	104
6.3.2	Theoretische Grundlagen	105
6.3.3	Implementierung	107
6.3.4	Ergebnisse für ein einfaches Stoffwechselmodell	109
6.3.5	Zusammenfassung	114
7	Zusammenfassung	117
	Literaturverzeichnis	121
	Anhang A Stoffwechselabschnitte	147
A.1	Glucose-Aufnahme	147
A.2	Glykolyse	148
A.3	Pentose-Phosphat-Weg	150
A.4	Synthese aromatischer Aminosäuren	151
	Anhang B Verfahrenstechnische Grundlagen	154
	Anhang C Datensätze	156
C.1	Experiment WT1	156
C.2	Experiment MU1	160
C.3	Experiment MU2	163
	Anhang D Enzymkinetiken	167
D.1	Allgemeines	167
D.2	Einfache Inhibitionskinetiken	168
D.3	Kooperativität allosterischer Enzyme	170
D.4	Kinetiken mit mehr als einem Substrat	170
	Anhang E Vereinfachungen des Reaktionsnetzwerkes	172
	Anhang F Simulation und Optimierung	174
F.1	Numerische Integration	174
F.1.1	Allgemeines	174
F.1.2	Einschrittverfahren	175

F.1.3	Absolute (lineare) Stabilität und steife Systeme	179
F.1.4	Mehrschrittverfahren	181
F.2	Mathematische Optimierung	184
F.2.1	Erklärung von Begriffen	184
F.2.2	Unrestringierte nichtlineare Optimierung	185
F.2.3	Nelder-Mead-Simplexmethode und Subplex-Algorithmus	187
Anhang G Ergänzungen zu den Stoffwechselmodellen		194
G.1	Parameterwerte der Modelle zur Beschreibung der AAA-Synthese	194
G.2	Zusammenfassung der Resultate für den <i>E.coli</i> -Wildstamm	195
G.3	Auflistung der Modelle für den <i>E.coli</i> -Wildstamm	202

Abbildungsverzeichnis

1–1	Beispiel für die gezielte Optimierung der Stoffwechselleistung von <i>C.glutamicum</i> zur Steigerung der Raum-Zeit-Ausbeute an L-Lysin	1
1–2	Konzept der Arbeit	3
2–1	Strukturelle und funktionelle Sichtweise einer Bakterienzelle	6
2–2	Relaxationszeiten unterschiedlicher Prozesse in Mikroorganismen	9
2–3	Modell-basierte Repräsentationen des metabolischen Netzwerkes	12
2–4	Modellierungs- und Validierungsprozeß	21
3–1	Fermentationsverlauf der Hauptkultur aus Experiment MU1	29
3–2	Trajektorien von PEP und G6P aus den Experimenten WT1, MU1 und MU2	30
3–3	Ablaufschema zur Erstellung des Computermodells	35
4–1	Bestimmung des Meßergebnisses für eine direkt gemessene Größe	43
4–2	Ermittlung des Meßergebnisses für eine nicht direkt gemessene Größe	44
4–3	Meßergebnisse und Zusammensetzung des Größtfehlers für zwei Metabolite (MU1)	47
4–4	Konzentrationsabhängigkeit des Größtfehlers	48
4–5	Startfenster des Auswertungswerkzeuges PulseSTAT	50
4–6	Glättung der G6P-Trajektorie aus Experiment WT1	52
4–7	Modellreaktion zur Erläuterung der Zeitreihenanalyse	56
4–8	Ergebnisse der Zeitreihenanalyse für eine lineare Reaktionssequenz	57
4–9	Flußschema zur Implementierung der Zeitreihenanalyse	59
4–10	Algorithmus zur agglomerativen hierarchischen Clusteranalyse	60
4–11	Ergebnisse der Zeitreihenanalyse für verrauschte Trajektorien	61
4–12	Ergebnisse der Zeitreihenanalyse (TSA) für manipulierte Trajektorien	62
4–13	Ergebnisse der TSA für das Modell aus (Chassagnole et al. 2003)	64
4–14	Ergebnisse der TSA für die Datensätze aus den Experimenten WT1 und MU1	66
5–1	Simulationsergebnisse des Modells 1 zur Beschreibung der Trajektorien aus Experiment MU1	74
5–2	Simulationsergebnisse des Modells 1 zur Beschreibung der Trajektorien aus Experiment MU2	74
5–3	Simulationsergebnisse der Modelle 4 und 5 zur Beschreibung der Trajektorien aus Experiment MU2	76
5–4	Modellfamilie zur Beschreibung der Aromatenbiosynthese in <i>E.coli</i> -Mutanten	77
5–5	Stoffwechselnetzwerk der Glykolyse und des PPP in einem <i>E.coli</i> -Wildstamm	78
5–6	Ergebnisse des Modells aus den Tabellen 5–3 und 5–4; simulierte Metabolittrajektorien	83
5–7	Ergebnisse des Modells aus den Tabellen 5–3 und 5–4; simulierte Flüsse	83

6–1	Implementation der Sensitivitätsberechnung in MMT	90
6–2	Sensitivitätsfunktionen bzgl. eines Parameter aus Modell 10 vor und nach der Modellreduktion	91
6–3	Simulationsergebnisse von Modell 10 vor und nach der Parameterreduktion	94
6–4	Sensitivitätsfunktionen bzgl. zweier korrelierender Parameter aus Modell 10	101
6–5	Schemata zur Implementierung des auf den Bootstrap-Datensätzen basierenden Anpassungstests	108
6–6	Makrokinetisches Stoffwechselmodell zur Untersuchung der Anwendbarkeit des Bootstrap-Verfahrens	110
6–7	Abgewandeltes makrokinetisches Stoffwechselmodell zur Untersuchung der Anwendbarkeit des Testverfahrens	111
7–1	Hauptergebnisse der vorliegenden Dissertation	117
A–1	PEP:Phosphotransferase System	147
A–2	Reaktionen der Glykolyse	149
A–3	Schema des Pentose-Phosphat-Weges	151
A–4	Schema der Synthese aromatischer Aminosäuren in <i>E.coli</i>	152
B–1	Wachstumskurve	154
C–1	Überblick über die in Experiment WT1 erfaßten Metabolite	156
C–2	Trajektorien aus Experiment WT1, Teil 1	157
C–3	Trajektorien aus Experiment WT1, Teil 2	158
C–4	Trajektorien aus Experiment WT1, Teil 3	159
C–5	Überblick über die in Experiment MU1 erfaßten Metabolite	160
C–6	Trajektorien aus Experiment MU1, Teil 1	161
C–7	Trajektorien aus Experiment MU1, Teil 2	162
C–8	Trajektorien aus Experiment MU1, Teil 3	163
C–9	Überblick über die in Experiment MU2 erfaßten Metabolite	163
C–10	Trajektorien aus Experiment MU2, Teil 1	164
C–11	Trajektorien aus Experiment MU2, Teil 2	165
C–12	Trajektorien aus Experiment MU2, Teil 3	166
D–1	Molekulare Reaktionsmechanismen von Enzymreaktionen	169
F–1	Stabilitätsgebiete der Euler-Cauchy-Verfahren	180
G–1	Teil 1 der Modelltopologien zur Beschreibung von <i>E.coli</i> K12	197
G–2	Teil 2 der Modelltopologien zur Beschreibung von <i>E.coli</i> K12	198
G–3	Ausgewählte Ergebnisse von Modell 37	199
G–4	Ausgewählte Ergebnisse von Modell 66	200
G–5	Ausgewählte Ergebnisse von Modell 69	201

Tabellenverzeichnis

3–1	Übersicht der verwendeten Mikroorganismen	28
4–1	Wert und geschätzte Meßunsicherheit der unmittelbaren Größen in Experiment MU1 .	46
4–2	Bestimmungs-, Nachweis- und Erfassungsgrenzen für die Kalibrierfunktionen aus Experiment MU1	49
4–3	Intrazelluläre Metabolitkonzentrationen im Glucose-limitierten Fließgleichgewichtszustand – ein Vergleich mit Literaturwerten	55
4–4	Datenmanipulationen zur Untersuchung der Eignung der TSA zur Datenvalidierung . .	62
5–1	Kinetische Gleichungen der Modelle 1 und 2 zur Beschreibung der Datensätze aus den Experimenten MU1 und MU2	73
5–2	Kinetische Gleichungen der Modelle 4 und 5 zur Beschreibung der Datensätze aus den Experimenten MU1 und MU2	75
5–3	DGL-System eines Modells zur Beschreibung des Zentralstoffwechsels in <i>E.coli</i> K12 .	79
5–4	Reaktionskinetische Terme für das Modell aus Tab. 5–3	80
5–4	Reaktionskinetische Terme für das Modell aus Tab. 5–3	81
5–4	Reaktionskinetische Terme für das Modell aus Tab. 5–3	82
5–5	Charakteristika der Modelle zur Beschreibung des Substrat-Puls-Experimentes WT1 im Vergleich mit veröffentlichten Modellen	85
6–1	Geschätzte Parameterwerte und korrespondierende Schätzfehler für Modell Nr. 10 vor und nach der Modellreduktion	94
6–1	Geschätzte Parameterwerte und korrespondierende Schätzfehler für Modell Nr. 10 vor und nach der Modellreduktion	95
6–1	Geschätzte Parameterwerte und korrespondierende Schätzfehler für Modell Nr. 10 vor und nach der Modellreduktion	96
6–1	Geschätzte Parameterwerte und korrespondierende Schätzfehler für Modell Nr. 10 vor und nach der Modellreduktion	97
6–2	Ergebnisse für die Parameterreduktion der Modelle aus Tab. 5–5	98
6–3	Ergebnisse des auf den Bootstrap-Datensätzen basierenden Anpassungstests für ein makrokinetisches Stoffwechselmodell	111
6–4	Vergleich verschiedener Kriterien zur Beurteilung der Anpassungsgüte für zwei makrokinetische Stoffwechselmodelle	112
E–1	Vereinfachungen des Reaktionsnetzwerkes	173
F–1	Allgemeines Butcher-Schema	176
F–2	Butcher-Schemata gebräuchlicher Runge-Kutta-Verfahren	176
F–3	Vektornormen	177
F–4	Adams-Verfahren	181
G–1	Parameterwerte der Modelle 1 und 2 aus Abschnitt 5.2	194

G-2	Parameterwerte der Modelle 3, 4 und 5 aus Abschnitt 5.2	195
G-3	Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1	202
G-3	Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1	203
G-3	Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1	204
G-3	Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1	205
G-3	Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1	206
G-3	Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1	207
G-3	Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1	208
G-4	Für die verschiedenen Reaktionsschritte verwendete Kinetiken in den Modellen der Topologien 0.a und 1	208
G-5	Für die verschiedenen Reaktionsschritte verwendete Kinetiken in den Modellen der Topologie 2	209
G-6	Für die verschiedenen Reaktionsschritte verwendete Kinetiken in den Modellen der Topologie 3	209
G-7	Für die verschiedenen Reaktionsschritte verwendete Kinetiken in den Modellen der Topologien 4 (links) und 5 (rechts)	210
G-8	Zuordnung der Modellnummern aus Tabellen G-4 bis G-7 zu den Namen bzw. der Nummern der Modelle im Hauptteil	210

Abkürzungsverzeichnis

Anm. : Die Abkürzungen der Enzyme entsprechen denen der korrespondierenden Gene. Sie besitzen aber große Anfangsbuchstaben, und ihre Schreibweise ist nicht kursiv.

Abkürzung	Bedeutung
2PG	2-Phosphoglycerat
3PG	3-Phosphoglycerat
6PG	6-Phosphogluconat
AAA	Biosyntheseweg der Aromatischen Aminosäuren
Abb.	Abbildung
AcCoA	Acetyl-CoenzymA
ADP	Adenosindiphosphat
<i>aldo</i>	Gen der Fructosebisphosphat-Aldolase
<i>amp^R</i>	Ampicillin-Resistenz-Gen
AMP	Adenosinmonophosphat
ATP	Adenosintriphosphat
<i>aroA</i>	Gen der 5-Enolpyruvoyl-3-phosphat-Synthase
<i>aroB</i>	Gen der 3-Dehydrochinat-Synthase
<i>aroD</i>	Gen der 3-Dehydrochinat-Dehydratase
<i>aroE</i>	Gen der Shikimat-Dehydrogenase
<i>aroF</i>	Gen der L-Tyrosin-spezifischen 3-Deoxy-D-arabino-heptulosonat-7-phosphat-Synthase
<i>aroG</i>	Gen der L-Phenylalanin-spezifischen 3-Deoxy-D-arabino-heptulosonat-7-phosphat-Synthase
<i>aroH</i>	Gen der L-Tryptophan-spezifischen 3-Deoxy-D-arabino-heptulosonat-7-phosphat-Synthase
<i>aroK</i>	Gen der Shikimat-Kinase I
<i>aroL</i>	Gen der Shikimat-Kinase II
ATP	Adenosintriphosphat
AWP	Anfangswertproblem
BDF	backward differentiation formulas
BG	Bestimmungsgrenze
BST	Biochemical Systems Theory
bzw.	beziehungsweise
C5P	Summenkonzentration der gemessenen Zucker-5-Phosphate des PPP
cAMP	cyclisches Adenosinmonophosphat
Cit	Citrat
CvM	Cramér-von-Mises-Teststatistik
DAH	3-Deoxy-D-arabino-heptulose
DAHP	3-Deoxy-D-arabino-heptulosonat-7-phosphat
DGL	Differentialgleichung
DGP	1,3-Diphosphoglycerat
DHAP	Dihydroxyacetonphosphat (Glyceronphosphat)
<i>dahps</i>	Summe der Gene aller DAHP-Synthasen (Modellgröße)
DHQ	3-Dehydrochinat

Abkürzung	Bedeutung
DHS	3-Dehydroshikimat
DIN	Deutsches Institut für Normung
DNS	Desoxyribonukleinsäure
E4P	Erythrose-4-phosphat
<i>E.coli</i>	<i>Escherichia coli</i>
EG	Erfassungsgrenze
EMP	Embden-Meyerhof-Parnas-Weg (Glykolyse)
<i>eno</i>	Gen der Enolase
EPSP	5-Enolpyruvoyl-3-phosphat
ESI-LC-MS	Elektrospray Ionisation Liquid Chromatography Mass Spectrometry
etc.	et cetera
F6P	Fructose-6-phosphat
FBP	Fructose-1,6-bisphosphat
<i>fbpase</i>	Gen der Fructose-1,6-bisphosphatase
fbr	feedback-resistent
FFT	Fast Fourier Transformation
G1P	Glucose-1-phosphat
G3P	Glycerol-3-phosphat
<i>g3pdh</i>	Gen der Glycerol-3-phosphat-Dehydrogenase
G6P	Glucose-6-phosphat
<i>g6pdh</i>	Gen der Glucose-6-phosphat-Dehydrogenase
GAP	Glyceraldehyd-3-phosphat
<i>gapdh</i>	Gen der Glyceraldehyd-3-phosphat-Dehydrogenase
GC-MS	Gas Chromatography Mass Spectrometry
Gluc	Glucose
HPLC	High Performance(Pressure) Liquid Chromatography
i.A.	im Allgemeinen
IBT	Institut für Biotechnologie
<i>in-silico</i>	am Rechner durchgeführte Experimente
<i>in-vitro</i>	Experimente unter künstlichen Bedingungen (im Reagenzglas)
<i>in-vivo</i>	Experimente im lebenden System
IPTG	Isopropylthio- β -D-galactosid
Kap.	Kapitel
K.S.	Kolmogorov-Smirnov-Teststatistik
LC-MS	Liquid Chromatography Mass Spectrometry
LSODE	Livermore Solver for Ordinary Differential Equations
MCA	Metabolic Control Analysis
MDS	Multidimensionale Skalierung
MFA	Metabolic Flux Analysis
MMT	Metabolic Modeling Tool
mRNS	Messenger(Boten)-RNS
MU1	Experiment mit <i>Escherichia coli</i> 4pF20 (L-Phenylalanin-Produzent)
MU2	Experiment mit <i>Escherichia coli</i> 4pF78 (L-Phenylalanin-Produzent)
MVP	Modellierungs- und Validierungsprozeß
NAD/NADH	Nicotinamid-adenin-dinucleotid (oxidierte/reduzierte Form)

Abkürzung	Bedeutung
NADP/NADPH	Nikotinamid-adenin-dinucleotid-phosphat (oxidierte/reduzierte Form)
NG	Nachweisgrenze
NMR	Nuclear Magnetic Resonance Spectrometry
NMS	Nelder-Mead-Simplex
OD ₆₅₀	Optische Dichte bei 650 nm
OxAc	Oxalacetat
<i>pdh</i>	Gen der Pyruvat-Dehydrogenase
PEP	Phosphoenolpyruvat
<i>pfk</i>	Gen der 6-Phosphofructokinase
<i>pgdh</i>	Gen der 6-Phosphogluconat-Dehydrogenase
<i>pghl</i>	Gen der 6-Phosphogluconat-Dehydratase
<i>pgi</i>	Gen der Phosphoglucoisomerase
<i>pgk</i>	Gen der Phosphoglycerat-Kinase
<i>pgm</i>	Gen der Phosphoglycerat-Mutase
Phe	L-Phenylalanin
<i>pheA</i>	Gen der Chorismatmutase / Prephenatdehydratase
pO ₂	Sauerstoffpartialdruck
PPP	Pentose-Phosphat-Weg
<i>pps</i>	Gen der Phosphoenolpyruvat-Synthase
<i>ppc</i>	Gen der Phosphoenolpyruvat-Carboxylase
<i>pts</i>	Gen des Phosphoenolpyruvat:Phosphotransferase-Systems
<i>pykA, pykF</i>	Gene der Isoenzyme der Pyruvat-Kinase
Pyr	Pyruvat
R5P	Ribose-5-phosphat
Ri5P	Ribulose-5-phosphat
RK	Runge-Kutta
RNS	Ribonukleinsäure
<i>rpe</i>	Gen der Ribulose-5-phosphat-Epimerase
<i>rpi</i>	Gen der Ribose-5-phosphat-Isomerase
<i>rppk</i>	Gen der Ribose-5-phosphat-pyrophospho-Kinase
rRNS	ribosomale RNS
s.	siehe
S3P	Shikimat-3-Phosphat
S7P	Sedo-heptulose-7-phosphat
SBML	Systems Biology Markup Language
<i>S.cerevisiae</i>	<i>Saccharomyces Cerevisiae</i> (Hefe)
Shik	Shikimat
<i>ta</i>	Gen der Transaldolase
Tab.	Tabelle
<i>tim</i>	Gen der Triosephosphat-Isomerase
<i>tka</i>	Gen der Transketolase A
<i>tkb</i>	Gen der Transketolase B
tRNS	Transfer-RNS
TSA	Time Series Analysis (Zeitreihenanalyse)
Tyr	L-Tyrosin

Abkürzung	Bedeutung
<i>tyrA</i>	Gen der Chorismatmutase / Prephenatdehydrogenase
u.U.	unter Umständen
VBA	Visual Basic for Application
vgl.	vergleiche
VV& T	Verifizieren, Validieren und Testen
wt	Wildtyp
WT1	Experiment mit <i>Escherichia coli</i> K12
X5P	Xylulose-5-Phosphat
XML	eXtended Markup Language
z.T.	zum Teil

Symbole

Anm.: Aus Gründen der Übersichtlichkeit wird auf die gängige Notation der chemischen Konzentrationen mit eckigen Klammern verzichtet und statt dessen in den Formeln nur die betreffende Abkürzung des Metaboliten bzw. Enzyms verwendet.

Symbol	Bedeutung
#	Anzahl von Entitäten
$\mathbf{11}^T$	Identitätsmatrix
A	virtueller Metabolit
AIK	Akaike-Informations-Kriterium
b	Zahl von Bootstrap-Datensätzen
B	virtueller Metabolit
$\mathbf{B}$	zentriertes inneres Matrixprodukt
$\mathbf{c}$	Vektor der simulierten intrazellulären Metabolitkonzentrationen
$\mathbf{c}_0$	Vektor mit den Anfangswerten der Konzentrationen
$\tilde{\mathbf{c}}$	Vektor der experimentell bestimmten intrazellulären Metabolitkonzentrationen
$\tilde{\mathbf{c}}^*$	Konzentrationsvektor eines verrauschten <i>in-silico</i> -Datensatzes
$\bar{\mathbf{c}}$	Vektor mit den zeitlich gemittelten experimentellen Konzentrationswerten
$\mathbf{c}_{ex}$	Vektor der extrazellulären Metabolitkonzentrationen
$\tilde{c}_{\text{Kal}}$	Analytkonzentration in den Kalibrierproben
$\tilde{c}_{\text{LCMS}}$	über die LC-MS-Messung bestimmte Metabolitkonzentration
C	virtueller Metabolit
$\mathbf{C}$	Meßfehler-Kovarianzmatrix
$C_{jk}^{\mathbf{c}_0}$	Konzentrationskontrollkoeffizient
$C_{jk}^{\mathbf{v}_0}$	Flußkontrollkoeffizient
$Cov_{jk}(\tau)$	Kreuzkovarianzfunktion
$\mathbf{C}_p$	Parameterschätzfehler-Kovarianzmatrix
d	Zahl der Datensätze
d_{jk}	euklidische Distanz
D	virtueller Metabolit
D	Verdünnungsrate
$\mathbf{D}$	Matrix der euklidischen Distanzen
$\mathbb{D}$	Definitionsbereich
e_c	quantitativ erfaßbarer und korrigierbarer Anteil des systematischen Fehlers einer Einzelmeßgröße
e_s	Restfehler des systematischen Fehleranteils einer Einzelmeßgröße
e_z	Anteil des zufälligen Fehlers einer Einzelmeßgröße
$e(\Delta p_g)$	Objektivfunktion zur Beurteilung der Veränderung der Simulationsergebnisse in Abhängigkeit von der Änderung eines Parameters p_g
$\mathbf{E}$	Vektor der Enzymkonzentrationen
$\mathbb{E}$	Erwartungswert

Symbol	Bedeutung
f_{abs}	absoluter Fehleranteil
f_{rel}	relativer Fehleranteil
F	Fehlerquadratsumme
$\hat{F}$	minimierte Fehlerquadratsumme
FIM	Fisher-Informationsmatrix
F_{in}	Zuflußrate
$F_{verd, Extr.}$	Verdünnungsfaktor der intrazellulären Metabolitkonzentrationen infolge der Zellextraktion
$F_{verd, LCMS}$	Verdünnungsfaktor der intrazellulären Metabolitkonzentrationen infolge der Probenvorbehandlung für die LC-MS-Bestimmung
g	Index für die Modellparameter
$\mathbf{g}$	Gradienten-Vektor
$\mathbf{h}$	Vektor mit den Wichtungskoeffizienten der Residuen bei der Berechnung der Fehlerquadratsumme
H_0	Nullhypothese
$\mathbf{H}$	Hesse-Matrix
i	Index der Zeitpunkte
$\mathbf{I}$	Einheitsmatrix
j	Index für die Metabolite
$\mathbf{J}_c$	Jacobi-Matrix
$\mathbf{J}_p$	parametrische Jacobi-Matrix
k	Index für die Metabolite
k_{cat}	Turnover-Zahl (Katalyse-Konstante)
K_E	Spezifisches Zellvolumen
K_{eq}	Gleichgewichtskonstante
K_i	Inhibitionskonstante
K_m	Michaelis-Menten-Konstante
$\mathcal{L}$	Wert der Likelihood-Funktion
m	Zahl der Metabolite
m_{St}	Steigung der Kalibriergeraden
$\mathcal{M}$	Modell
n	Zahl der Zeitpunkte
$\mathbf{N}$	stöchiometrische Matrix
$N(0, \sigma^2)$	Normalverteilung
$os(p_g)$	durchschnittliche Sensitivität
$\mathbf{p}$	Parametervektor
$\hat{\mathbf{p}}$	Vektor der optimierten Parameterwerte
pr	Zahl der Kalibrierproben
q	Zahl der Modellparameter
q_s	Glucose-Aufnahmerate
Q	Gütefunktional der verallgemeinerten Regression
r	Index für die Reaktionsgeschwindigkeiten
R_∞	Grenzprozeß der Residuen
$R_{jk}(\tau)$	Kreuzkorrelationsfunktionen
R_n	empirischer Prozeß der Residuen

Symbol	Bedeutung
$\mathbb{R}$	Bereich der reellen Zahlen
$s_{x,y}$	Reststandardabweichung
$\tilde{s}$	normierte Parametersensitivität
$\bar{s}$	Stichproben-Standardabweichung
S	Matrix der Parametersensitivitäten
t	Faktor der „Student“ t-Verteilung
t	Zeit
T	transponierte Matrix
u	Meßunsicherheit
u	Vektor mit den Eingangssignalen eines Modells
ug	untere Grenze
$U(a, b)$	Gleichverteilung
v	Vektor der Reaktionsgeschwindigkeiten (Flüsse)
$V_{\max}$	Maximale Aktivität eines Enzyms (maximale Reaktionsgeschwindigkeit)
V_E	bei der Zellextraktion zugegebenes Volumen an Perchlorsäure und Resuspensionslösung
V_N	zur Neutralisation der Zellextrakte zugegebenes Volumen
V_{Probe}	Volumen der durch die Extraktion und Neutralisation verdünnten Probenlösung
V_{Puffer}	Volumen der für die LC-MS-Bestimmung zugegebenen Pufferlösung
V_S	dem Fermenter entnommenes Probenvolumen
V_X	Interzelluläres Flüssigkeitsvolumen, das nach der Extraktion und anschließenden Zentrifugation im Zellpellet verbleibt
V_R	Reaktorvolumen
w	Vektor des den simulierten Metabolittrajektorien überlagerten Rauschens
w_m	Zahl der Wiederholungsmessungen
$\tilde{W}$	Signalfläche einer LC-MS-Bestimmung
$\tilde{\mathbf{x}}$	Vektor der experimentell bestimmten unabhängigen Größen (Einzelmeßgrößen)
$\mathbf{x}^*$	unabhängige Größen eines Bootstrap-Datensatzes
$\tilde{x}_c$	wahrscheinlichster Wert für den wahren Wert der Einzelmeßgröße
$\tilde{x}_w$	wahrer Wert einer Einzelmeßgröße
X_1	virtueller Metabolit
X_2	virtueller Metabolit
y	Vektor mit den simulierten Trajektorien der abhängigen Modellgrößen
$\mathbf{y}^*$	abhängige Größen eines Bootstrap-Datensatzes
$\tilde{\mathbf{y}}$	Vektor der experimentell bestimmten Zustandsvariablen des Systems
$Y_{x/s}$	Ausbeutekoeffizient
z	Anzahl der Reaktionsgeschwindigkeiten im Modell
α	Signifikanzniveau

Symbol	Bedeutung
γ	Eigenvektor
Γ	Matrix der Eigenvektoren
δ	infinitesimal kleine Änderung
δ_{jk}	Kronecker-Delta
Δ	Meßunsicherheit
Δ	Laplace-Operator
ϵ	Residuum
ϵ^0	Elastizität
Θ_{90}	Mischzeit
λ	Eigenwert
Λ	Matrix der Eigenwerte
μ_{max}	Wachstumsrate
ν	Index Zeitverschiebung bei der Zeitreihenanalyse
ρ	Korrelationskoeffizient
σ	Standardabweichung
σ^2	Varianz
$\Sigma(t_i)$	Varianz-Kovarianz-Matrix der Residuen zum Zeitpunkt t_i
τ	Zeitverschiebung bei der Zeitreihenanalyse

1 Einleitung und Zielstellung

1.1 Einleitung

Modelle helfen, Fragestellungen zu beantworten, indem sie erlauben, die Ausschnitte der Realität, aus denen sie stammen, zu simulieren. Minsky (1968) setzt somit ein Modell zu einem konkreten Tupel aus System und Experiment in Beziehung: „ein Modell (M) für ein System (S) und ein Experiment (E) ist etwas, auf das E angewendet werden kann, um Fragen über S zu beantworten“.

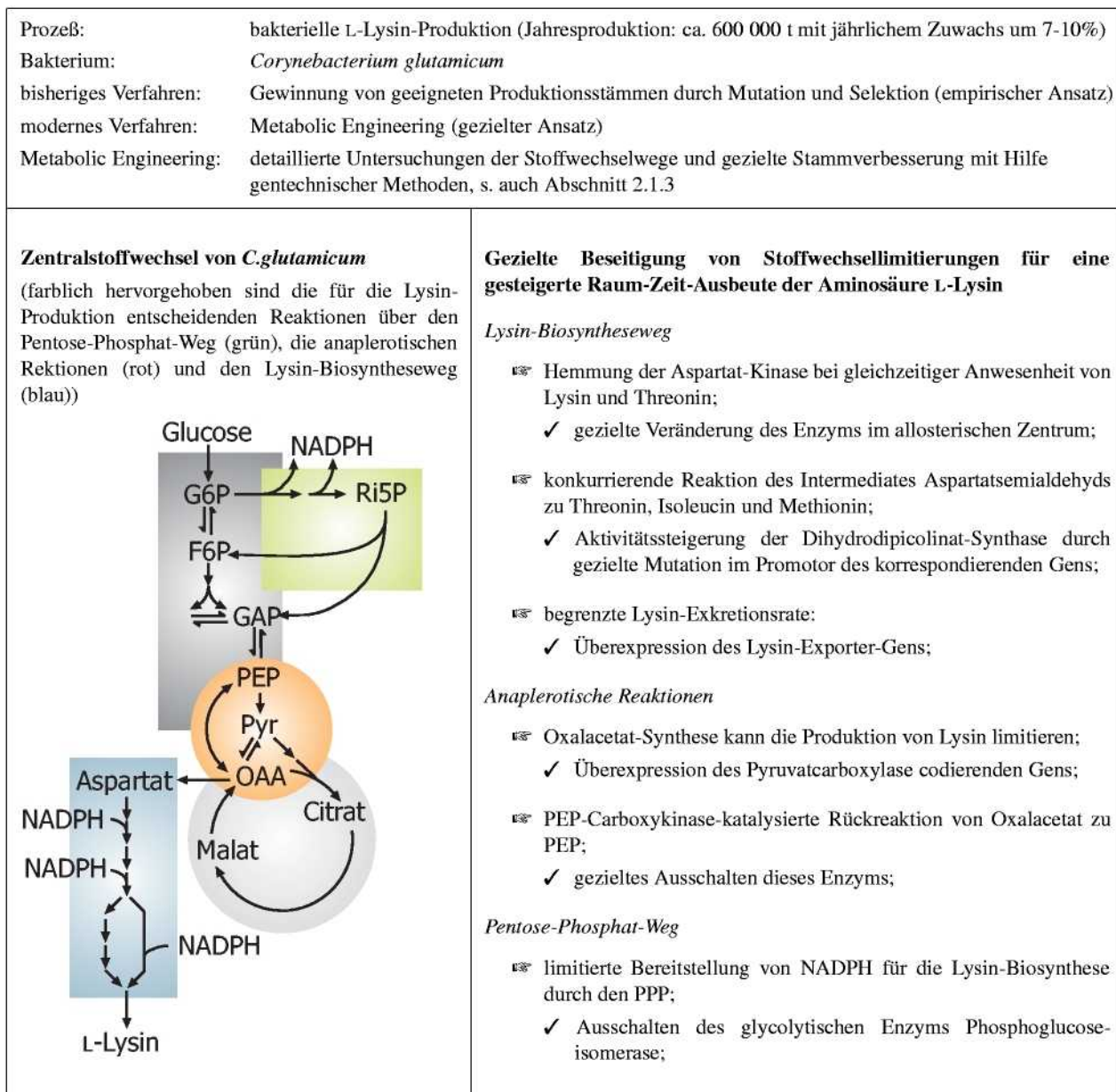


Abb. 1–1 Beispiel für die gezielte Optimierung der Stoffwechsellleistung von *C. glutamicum* zur Steigerung der Raum-Zeit-Ausbeute von L-Lysin, s. (Kennerknecht et al. 2003)

Modelle in der Biotechnologie bleiben dieser Definition treu. Eine ihrer zentralen Fragestellungen ist laut Bailey (1998): „darüber nachzudenken (und zu berechnen), welche Komponenten und Interaktionen in einem komplexen System wichtig sind“. In diesem Sinne wird für die Beschreibung der Vorgänge in

einer mikrobiellen Zelle das Bild einer Fabrik erschaffen, in der biotechnologisch relevante Produkte über Maschinerien, wie Enzyme, Ribosomen etc., gefertigt werden (Giuseppin et al. 1999). Wenn es gelingt, experimentelle Daten über einen bestimmten Produktionsprozeß zu erfassen und diese mit einem Modell zu beschreiben, sollte es möglich sein, wesentliche Charakteristika des Prozesses quantitativ zu erschließen. Insofern können auf der Basis des Modells optimale Fermentationsbedingungen als auch gezielte genetische Veränderungen der Zellen vorhergesagt werden, um z.B. die Raum-Zeit-Ausbeute der gewünschten Produkte zu steigern (Gombert und Nielsen 2000). Ein Beispiel für die erfolgreiche Anwendung von detaillierten Kenntnissen über den mikrobiellen Stoffwechsel wird in Abb. 1–1 demonstriert.

Das Potential moderner High-Throughput-Technologien, außerordentlich große Mengen biologischer Daten zu gewinnen, läßt darüber hinaus sogar die Vision einer *in-silico*-Zelle als Abbildung des natürlichen Systems entstehen (Westerhoff 2001; Westerhoff et al. 2003). Sie stellt jedoch kein Modell im eigentlichen Sinne mehr dar, da das System ohne zusätzliche vereinfachende Annahmen vollständig durch die Daten beschrieben werden kann. Im Gegensatz dazu ist Modellierung als ein Prozeß der Abstraktion zu verstehen, der aufgrund der verfolgten Fragestellung oder allein schon wegen fehlender Daten Vereinfachungen und die Vernachlässigungen von Details umfaßt. Daher kann ein Modell für ein Experiment gültige Aussagen machen und für ein anderes nicht. Nicht nur aus diesem Grund ist spätestens vor der Verwendung des Modells seine Gültigkeit im Rahmen der Validierung zu prüfen. Ohne Validierung bleibt die Untersuchung von Modellen *l'art pour l'art*.

1.2 Zielstellung

Ziel der vorliegenden Arbeit ist es, valide Differentialgleichungsmodelle für die Beschreibung mikrobieller Stoffwechselvorgänge zu entwickeln. Ausgangspunkt bilden Datensätze aus Substrat-Puls-Experimenten, die in Kapitel 3.2 näher beschrieben werden. Die Arbeit ordnet sich in das langfristige Projekt am Institut für Biotechnologie II des Forschungszentrums Jülich ein, auf diesem Bereich eine Rückkopplung zwischen Modellen und Experimenten im Sinne der Versuchsplanung zu erreichen. Für diese Arbeit stand daher der Prozeß zur Modellbildung und Validierung und weniger ein konkretes Modell, aus dem Aussagen abgeleitet werden, im Mittelpunkt der Untersuchungen.

Die für das Projekt notwendigen Voraussetzungen wurden in verschiedenen Vorarbeiten geschaffen:

- Datensätze: der künftig mit WT1 bezeichnete Datensatz aus einem Experiment mit dem Bakterium *Escherichia coli* K12 lag bereits zu Beginn der Arbeit vor (Buchholz 2001); zwei weitere Datensätze (MU1 und MU2) aus Experimenten mit genetisch modifizierten *E.coli*-Stämmen sind im Mai 2003 entstanden (Oldiges 2004);
- Modellierungswerkzeug: zur Erstellung und Simulation der Modelle wurde das Modellierungswerkzeug MMT, Metabolic Modeling Tool, (Hurlebaus 2001) bzw. ab Januar 2003 die Software MMT2 (Haunschild 2001) genutzt;
- Vorwissen: die Informationen über das zugrundeliegende Stoffwechselnetzwerk und die Enzymkinetiken sind aus Online-Datenbanken, wie Kegg (Kanehisa und Goto 2000), Brenda (Schomburg et al. 2002), ExPASy (Appel et al. 1994), EcoCyc (Karp et al. 2000), erhältlich;

Die wesentlichen, den Modellbildungs und Validierungsprozeß bestimmenden Fragestellungen sind in Abb. 1–2 zusammengefaßt. Um entscheiden zu können, welche Daten in das Modell eingehen sollen, war die Datenqualität zu beurteilen. Da die Modellierungswerkzeuge im Rahmen dieser Arbeit zum ersten Mal in größerem Umfang verwendet wurden, galt es, ihre Funktionsfähigkeit, d.h. die Korrektheit der Simulationsergebnisse und die Betriebsstabilität, zu untersuchen. Für die Erstellung der Modelle

waren die geeigneten Informationen aus zahlreichen und teilweise widersprüchlichen Angaben in den Datenbanken zu selektieren. Da auf diesem Weg mehrere Modelle gefunden werden, die die Meßdaten angemessen wiedergeben, sollten sie anhand von weiteren Kriterien unterschieden werden. Schließlich war zu prüfen, welche Aussagen aus den resultierenden Modellen abgeleitet werden können. Zur Beantwortung der Fragen waren geeignete Methoden auszuwählen und ihr Potential bzw. ihre Grenzen für diesen Zweck zu evaluieren. Dies wird in den folgenden Kapiteln erläutert.

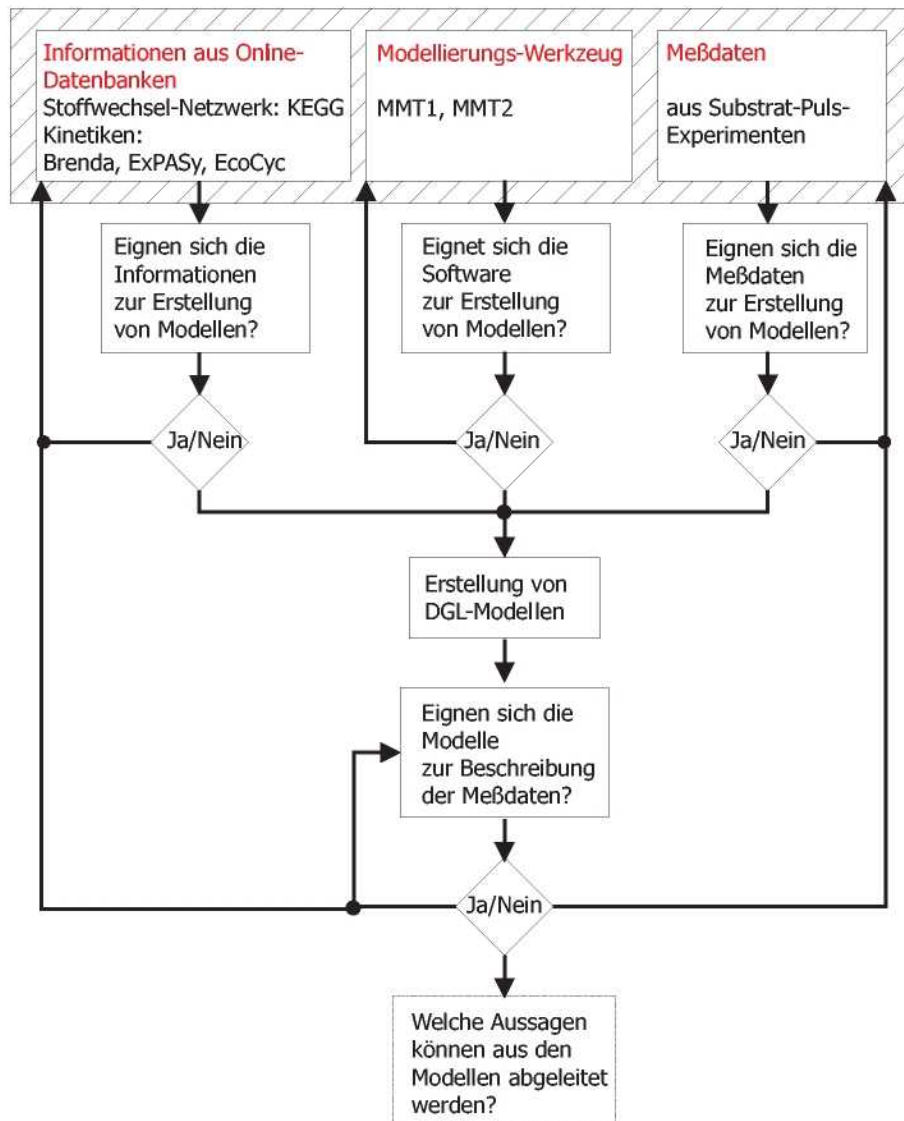


Abb. 1–2 Konzept der Arbeit: wesentliche Fragestellungen des Modellbildungs und Validierungsprozeß

1.3 Gliederung der Arbeit

Die Grundlagen, die für das Verständnis der Arbeit notwendig sind, werden in Kapitel 2 bzw. in Anhang A bis Anhang F vorgestellt^{1.1}. Im darauffolgenden Kapitel wird die Wahl der Validierungsmethoden begründet und das daraus resultierende Konzept des Modellierungs- und Validierungsprozeß dargestellt. Die Ergebnisse für die Implementierung, Evaluation und Anwendung der einzelnen Methoden werden in den Kapiteln 4, 5 und 6 diskutiert. Dabei wird das Kapitel 5, das sich auf die Modellbildung

^{1.1} Aufgrund des interdisziplinären Charakters der Arbeit sind diese relativ umfangreich.

bezieht, durch Anhang G ergänzt, in dem ein Großteil der im Rahmen dieser Arbeit erstellten Modelle detailliert aufgelistet ist. Kapitel 7 faßt die wesentlichen Ergebnisse der Arbeit zusammen.

2 Theoretische Grundlagen

2.1 Biologische Grundlagen

2.1.1 Beschreibung einer prokaryotischen Zelle

„Mikroorganismus“ oder „Mikrobe“ ist ein Sammelbegriff für Kleinlebewesen, die nur schwer oder gar nicht mit dem bloßen Auge erkennbar sind. Zu den Mikroben gehören bestimmte Eukaryoten, wie z.B. einige Abteilungen der Pilze, und Prokaryoten, d.h. Archae- und Eubakterien. Die Prokaryoten unterscheiden sich von den Eukaryoten in erster Linie durch das Fehlen eines vollständigen, durch eine Membran eingehüllten Zellkerns. Gemeinsame Strukturmerkmale der Prokaryoten sind in Abbildung 2–1 links gezeigt, s. (Lengeler et al. 1999). Darüber hinaus weisen Bakterien jedoch auch gruppenspezifische Charakteristika auf.

Die einzelnen Organellen der Bakterienzelle befinden sich in einer gelartigen Grundsubstanz innerhalb der Zelle, dem Cytoplasma. Das Cytoplasma wird durch die Cytoplasmamembran eingeschlossen. In dieser sind neben Enzymen, die für den Aufbau der Zellwand notwendig sind, selektive Transportproteine für einen kontrollierten Stoffaustausch und Enzyme der Atmungskette eingebettet. Mit Ausnahme der Mykoplasmen besitzen Bakterien eine Zellwand, deren Hauptbestandteil Murein ist. Bei gram-negativen Bakterien^{2.1} findet sich die Mureinschicht unmittelbar an die Cytoplasmamembran angrenzend und ist über Lipoproteine mit einer zweiten äußeren Lipid-Doppelmembran verbunden. Zwischen Mureinschicht und äußerer Membran ist ein Zwischenraum ausgebildet, der als periplasmatischer Raum bezeichnet wird. Gram-positive Bakterien besitzen keine Außenmembran und die Mureinschicht, die die Plasmamembran umgibt, ist ungefähr dreimal dicker als die der gram-negativen Bakterien.

Geißeln oder Flagellen in unterschiedlicher Anordnung –monotriche (einzelne), lophotriche (büschelartige) oder peritriche (auf der gesamten Oberfläche) verteilte Geißeln– ermöglichen beweglichen Bakterien die Fortbewegung. Kleinere fädige Glykokalix-Strukturen, die in Form von Fimbrien (Fransen) oder Pili (Haare) die Oberfläche der Bakterien bedecken, dienen dazu, sich auf der Unterlage fest zu heften.

Im Cytoplasma befindet sich ein einziges, Desoxyribonucleinsäure-, bzw. DNS-Molekül, das Nukleid. Bei manchen Bakterien gibt es außer dem Chromosom ein oder mehrere weitere, ringförmige DNS-Moleküle in der Zelle. Diese Plasmide sind vom Chromosom völlig unabhängig und für die Funktion oder Reproduktion der Zelle nicht unbedingt notwendig. Weiterhin beinhaltet das Cytoplasma einer Zelle im Wesentlichen 10-30.000 Ribosomen zur Proteinsynthese, mehrere Tausend Proteine, zahlreiche Stoffwechselprodukte und Kofaktoren sowie eine Reihe anorganischer Ionen (Schlegel 1992).

Es gibt in der Bakterienzelle eine einfache Arbeitsteilung, s. Abb. 2–1 rechts (Lengeler et al. 1999). Das Nukleid speichert und übermittelt die genetische Information zur Proteinsynthese. Während der Transkription wird die genetische Information des doppelsträngigen Nukleids von einem Enzymsystem, den Ribonukleinsäure-, bzw. RNS-Polymerasen, auf einen RNS-Strang übertragen, dessen Basensequenz zu einem der DNS-Stränge komplementär ist. Es werden drei verschiedene Arten von RNS gebildet. Die Boten- oder Messenger-RNS (mRNS) trägt die Sequenzen, die eine oder mehrere Polypeptide kodieren. Die Moleküle der ribosomalen RNS (rRNS) sind mit Proteinen zu den Ribosomen, den Funktionseinheiten der Proteinsynthese verbunden. Die Transfer-RNS (tRNS) liest während der an den Ribosomen stattfindenden Translation die in der mRNS kodierte Information ab und überträgt eine

^{2.1}Die 1884 von Gram entwickelte Gram-Färbemethode differenziert aufgrund des unterschiedlichen Zellwandaufbaus zwischen gram-positiven und gram-negativen Bakterien: der im Cytoplasma gebildete violette Komplex mit dem Farbstoff Gentiana- oder Kristallviolett kann bei gram-negativen Bakterien leicht ausgewaschen werden, während die Zellwand gram-positiver Bakterien gefärbt bleibt.

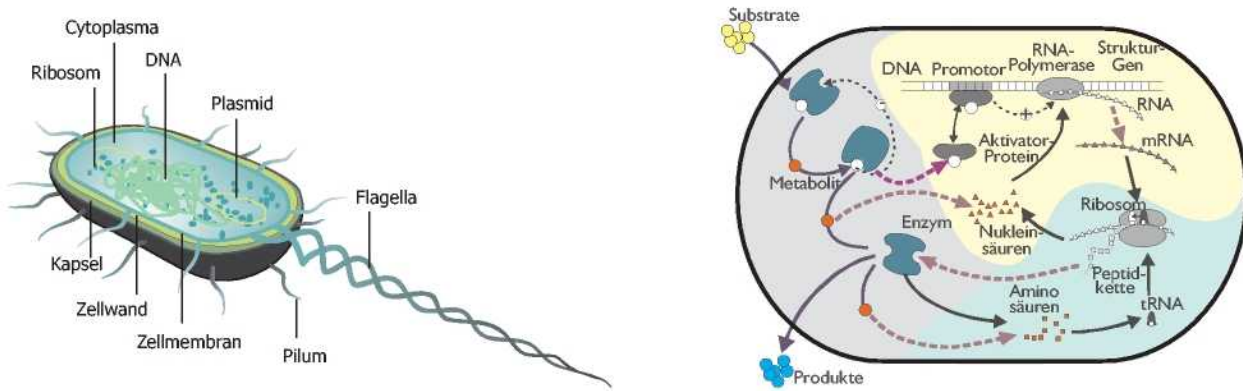


Abb. 2-1 Strukturelle (links) und funktionelle Sichtweise einer *E.coli*-Zelle (rechts); gestrichelte Pfeile: die Substanz an der Pfeilspitze ist Bestandteil der Substanzklasse am Pfeilende; durchgehende Pfeile: chemische Reaktion

passende Aminosäure auf die wachsende Polypeptidkette der Proteine.

Da die Proteinsynthese einen hohen Energieaufwand erfordert, gibt es zahlreiche Mechanismen, um die Proteinmenge in der Zelle zu regulieren (Lehninger et al. 1994). Die Transkription verläuft z.B. selektiv, d.h. es wird nur die genetische Information transkribiert, die die Zelle unter den gegebenen physiologischen Bedingungen benötigt. Hierzu gehen dem zu transkribierenden DNS-Strang bzw. den Strukturgenen Steuersequenzen, die Promotoren, voraus, an denen die Transkription initiiert wird und die Regulatoren der Transkription angreifen können^{2.2}. Diese Aktivierung von Genen zur Anfertigung der zugehörigen Proteine wird als Expression bezeichnet. Einige Proteine werden ständig in der Zelle benötigt und ihre Gene, die sogenannten Haushaltsgene, werden daher auf einem nahezu gleichbleibenden Niveau exprimiert bzw. gebildet, was als konstitutive Genexpression bezeichnet wird. Im Gegensatz dazu gibt es Genprodukte, deren Konzentration unter ganz bestimmten molekularen Bedingungen steigt bzw. fällt. Sie werden als induzierbar bzw. repressibel bezeichnet, und der Vorgang, durch den die Genexpression erhöht bzw. gesenkt wird, heißt Induktion bzw. Repression (Lewin 2000).

Die Aufgaben der Proteine bzw. Eiweiße sind vielfältig (Alberts et al. 2002). So gibt es u.a. Strukturproteine (z.B. Collagen), Transportproteine (z.B. Hämoglobin), Immunproteine oder Hormonproteine. Die Enzyme unter ihnen katalysieren die Reaktionen des Stoffwechsels. Die Intermediate der von den Enzymen katalysierten Reaktionskaskaden können die Aktivität der Enzyme als Aktivatoren steigern oder als Inhibitoren senken. Diese sogenannte allosterische Regulation ermöglicht es den Mikroorganismen, sich an rasch ändernde Umgebungsbedingungen anzupassen.

Der Stoffwechsel oder Metabolismus als Gesamtheit aller in der Zelle stattfindenden Reaktionen kann in den Katabolismus und den Anabolismus unterteilt werden (Wilson und Hunt 1994; Madigan et al. 2002). Unter Katabolismus werden alle Energie- und Reduktionsäquivalente erzeugenden Reaktionen verstanden, wobei nicht in jedem Fall komplexe organische Moleküle abgebaut werden müssen. Es gibt für die Mikroorganismen verschiedene, alternative Möglichkeiten, Energie zu gewinnen. Bei der Fermentation werden Elektronen von einer stark reduzierten organischen Verbindung auf ein organisches Oxidationsmittel übertragen. Der Energiegewinn ist gering, da die Produkte nicht vollständig oxidiert werden. Mehr Energie kann bei der aeroben Respiration mit Sauerstoff als endgültigem Elektronenakzeptor gewonnen werden. Beispiele für die aerobe Respiration sind der Abbau von Glucose über die Glykolyse, den Pentose-Phosphat-Weg (PPP) und den Citronensäurezyklus zu Kohlendioxid und Wasser, s. Anhang A. Die reduzierte Form des Nicotinamid-adenin-dinucleotid (NADH) dient als primärer Elektronendonator, der über die sich in der Zellmembran befindende Atmungskette unter Energiegewinn, d.h. Aufbau von Adenosintriphosphat (ATP), durch Sauerstoff oxidiert wird. Andere

^{2.2}Die gesamte Funktionseinheit aus Operatoren, wie den Promotoren, und den Strukturgenen wird als Operon bezeichnet.

anorganische Moleküle als Sauerstoff, wie z.B. Nitrat-, Sulfat oder Karbonat-Anionen, können bei der anaeroben Respiration als Elektronenakzeptoren dienen und zu H_2S , NH_3 oder Methan reduziert werden. Bei der Photosynthese wird Licht über Photophosphorylierungsreaktionen in Energie umgewandelt, wobei komplexe organische Moleküle häufig aus Kohlendioxid aufgebaut werden.

Die vom Katabolismus erzeugten Energie- und Reduktionsäquivalente werden vom Anabolismus benötigt, um komplexere organische Verbindungen aus einfacheren Vorläufermolekülen synthetisieren. Während prototrophe Stämme in der Lage sind, selbst alle benötigten Metabolite über anabole Reaktionen aus einer Kohlenstoffquelle und Mineralsalzen herzustellen, erstreckt sich der Nährstoffbedarf auxotropher Stämme über zusätzliche organische Wachstumsfaktoren, wie z.B. Aminosäuren, Nukleotide oder Vitamine. Aus den Ausgangssubstraten, die in Energie verbrauchenden Transportvorgängen in die Zelle gelangen oder Zwischenprodukte kataboler Stoffwechselwege darstellen, werden zunächst Monomere, d.h. Aminosäuren^{2,3} Nucleotide, Polysaccheride, Fettsäuren und Vitamine, aufgebaut. In Polymerisationsreaktionen entstehen aus den Monomeren Makromoleküle, wie Proteine, Nucleinsäuren, Peptidoglykane und Lipide. Die Polymere aggregieren zu höheren Organisationseinheiten, die wichtige Strukturen, u.a. Ribosomen, Chromosomen, Membranen und Zellwände, innerhalb der Zelle bilden. Hat die Zelle über anabole Reaktionen die lebensnotwendigen Stoffe in ausreichender Menge gebildet, beginnt sie sich zu teilen, indem das Chromosom repliziert und eine neue Zellwand eingezogen wird.

2.1.2 Der Mikroorganismus *Escherichia coli*

E.coli^{2,4} ist ein Vertreter einer großen Bakterienfamilie, der Enterobacteriaceae, die als gram-negative, stäbchenförmige, opportunistische oder pathogene, fakultative Anaerobier den menschlichen und tierischen Darmtrakt bewohnen. Fakultative Anaerobier können in der An- und Abwesenheit von Sauerstoff leben. Somit kann *E.coli* sowohl über die aerobe als auch die anaerobe Respiration (Nitrat-Atmung) und die Fermentation Energie gewinnen, wobei im letzteren Fall ein charakteristisches Gemisch aus Säuren und Gasen entsteht (Ingraham et al. 1983). Zum Wachstum reicht eine einzige Kohlenstoffquelle, die Glucose, neben einer Stickstoffquelle und Mineralsalzen aus. Unter optimalen Bedingungen (Komplexmedium, 37°C, pH-Wert zwischen 6,5-7,3) liegt die Generationszeit des ca. 1 - 4 μm großen Organismus bei 20 Minuten (Schlegel 1992).

E.coli reagiert auf Signale aus der Umgebung, wie Temperatur- oder pH-Änderungen und Chemikalien etc. Es bemerkt z.B. die Anwesenheit von Chemikalien und Gasen in seiner Umgebung und kann aufgrund seiner peritrichen Begeißelung auf diese zu bzw. von ihnen weg schwimmen. In einer Reaktion auf Änderungen der Temperatur oder der Osmolarität kann das Bakterium den Porendurchmesser der Porine in der äußeren Membran adaptieren, um größere (Nährstoff)-Moleküle aufzunehmen oder Hemmstoffe auszuschließen.

E.coli eignet sich aus verschiedenen Gründen als ein idealer Modellorganismus zur Klärung grundlegender Fragen der zellulären Biologie, z.B. auch zur Untersuchung von Stoffwechselwegen und ihren Regulationsmechanismen, wie in dieser Arbeit. Als Bakterium besitzt es generelle Vorteile, d.h. es ist einzellig, haploid (d.h. es beinhaltet nur ein Chromosom), aus der Gensequenz läßt sich im Wesentlichen direkt die Proteinsequenz ableiten, die Kultivierung ist preiswert, seine Generationszeit ist kurz, und eine große Zahl von Organismen lassen sich auf einem kleinen Raum halten. Darüber hinaus ist *E.coli* einer der sowohl molekularbiologisch als auch genetisch am detailliertesten untersuchten Organismen, s. (Blattner et al. 1997) für die komplette Sequenzierung des Genoms. Eine große Zahl von Datenbanken im Internet machen diese Informationen zugänglich (Appel et al. 1994; Kanehisa und Goto 2000; Karp et al. 2000; Schomburg et al. 2002). In der Biotechnologie wurde *E.coli* frühzeitig zur Darstellung von

^{2,3}In Anhang A ist die Synthese aromatischer Aminosäuren detailliert beschrieben, die im Rahmen dieser Arbeit u.a. modelliert wird.

^{2,4}*E.coli* wurde erstmalig von Theodor Escherich als „Bacterium coli commune“ charakterisiert (Escherich 1885).

Produkten wie Asparaginase, Penicillin-Acylase, Asparagin und Tryptophan (Varma et al. 1993) eingesetzt. Aufgrund der Verwendung mikrobieller Klonierungssysteme dient *E.coli* heute hauptsächlich als Wirtsorganismus zur Expression heterologer Proteine, z.B. Insulin, Interferon und Somatostatin (Swartz 1999) sowie zur Darstellung von Aminosäuren, Vitaminen oder Nukleotiden (Clark-Curtiss und Curtiss III 1983), und besitzt somit eine große wirtschaftliche Relevanz.

2.1.3 Metabolic Engineering

Unter „Metabolic Engineering“ versteht man den Einsatz von Methoden der DNS-Rekombinationstechnik, um die Produktion gewünschter Zelleigenschaften, Genprodukte und Metabolite zu regulieren bzw. gezielt zu steigern oder neuartige Genprodukte und Metabolite herzustellen (Stephanopoulos et al. 1998).

Bei der DNS-Klonierung wird ein spezifisches Gen oder DNS-Segment, das z.B. ein Genprodukt mit einer höheren Rate exprimieren soll, aus einem größeren Chromosom abgetrennt, mit einem kleineren DNS-Trägermolekül, dem Klonierungsvektor, verbunden und diese rekombinante DNS tausend- bis zu millionenfach repliziert. Als Klonierungsvektoren können z.B. Plasmide verwendet werden, die über den Vorgang der Transformation in Bakterienzellen eingeschleust werden. Um die erfolgreich transformierten Zellen zu selektieren, werden zusätzlich ein oder mehrere als Selektionsmarker bezeichnete Gene in den Vektor eingebaut, die die Zelle zum Wachstum unter bestimmten Bedingungen benötigt. In den Transformanten werden die Plasmide über Replikation vermehrt und auf einem Niveau von ca. 10-20 Kopien pro Zelle gehalten. Die entstandenen Mutanten unterscheiden sich aufgrund der geänderten DNS-Sequenz und evtl. des geänderten Phänotyps von einem willkürlich als Richtwert festgelegten Wildtypstamm. Häufig werden begleitend zur Transformation bestimmter Gene die entsprechenden Gene auf dem Chromosom deletiert, d.h. gezielt entfernt. Ohne darauf folgende Transformation werden die sogenannten „Knock-Out“-Mutanten auch zur funktionalen Analyse der Gene eingesetzt: eine Akkumulation bestimmter Metabolite in den Mutanten kann es ermöglichen, das deletierte Gen in Beziehung zu einem bestimmten Reaktionsschritt im biochemischen Netzwerk und dem zugehörigen Stoffwechselabschnitt zu setzen. Umgekehrt können über die Anhäufung bestimmter Metabolite in der Zelle Gene bzw. Genprodukte, die aufgrund einer relativ geringen Expressionsrate die Weiterverarbeitung dieser Metabolite limitieren, identifiziert und über gentechnische Eingriffe eliminiert werden.

2.2 Modellierung der Zelle

2.2.1 Allgemeines

In den vergangenen drei Jahrzehnten erzielten die Biowissenschaften hauptsächlich Fortschritte bei der Identifizierung und Analyse einzelner Bestandteile biologischer Systeme, so daß eine Fülle von Informationen über die verschiedenen Organisationseinheiten des Lebens (Organe, Gewebe, Zellen, etc.) existiert. Im Sinne dieses „Reduktionismus“ beschränkten sich bisher meist auch experimentelle und theoretische Untersuchungen auf der Zellebene darauf, das biologische Gesamtsystem in feinere Detailsysteme zu zerlegen, deren Einzelbestandteile dann jeweils möglichst genau beschrieben werden sollten. Eine Begründung für diese Vorgehensweise kann vor allem darin gefunden werden, daß die einzelnen sich in der Zelle abspielenden Vorgänge, s. Abschnitt 2.1.1, auf unterschiedlichen Zeitskalen stattfinden und daher getrennt voneinander betrachtet werden können, s. Abb. 2–2.

Für die Analyse von Genomen hat sich international der Begriff „Genomics“ etabliert, unter dem gängige, kommerzialisierte, zum Großteil automatisierte Techniken zur Sequenzierung des Genoms verstanden werden (Venter 2001; Brown 2002). Die erhaltenen Daten liefern jedoch noch keine Angaben über

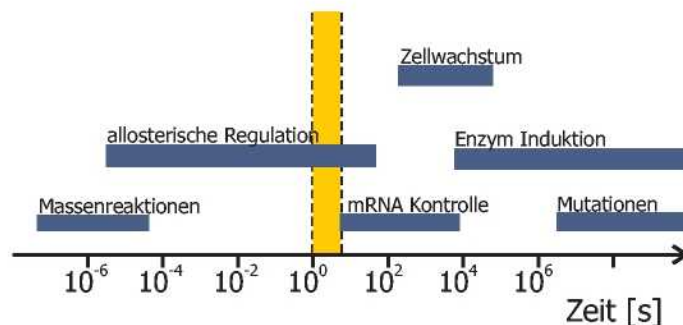


Abb. 2–2 Relaxationszeiten unterschiedlicher Prozesse in Mikroorganismen (adaptiert aus (Roels 1983), (Stephanopoulos et al. 1998) und (Fell 2003)); der durch gestrichelte Linien eingegrenzte Bereich kennzeichnet das Zeitfenster der Substrat-Puls-Experimente

die Aktivierung und Expression spezifischer Gene und werden daher vorwiegend für den Vergleich verschiedener Organismen bzw. zur Aufklärung phylogenetischer Beziehungen genutzt, s. z.B. (Gil et al. 2003; Richer et al. 2003). Als nächster Schritt schließt sich die Analyse der Genprodukte, d.h. der mRNS und der Proteine an. Die Gesamtheit dieser Arbeiten wird unter dem Namen „Functional Genomics“ zusammengefaßt und kann in drei Bereiche aufgeteilt werden:

- Die Transkriptomanalyse („Transcriptomics“) beschäftigt sich mit der Bestimmung von Genen, die unter bestimmten experimentellen Bedingungen exprimiert werden. Dazu werden mRNS-Transkriptionsprofile, d.h. mRNS-Konzentrationen verschiedener Stämme bzw. eines Stammes unter verschiedenen physiologischen Bedingungen, mit der DNS-Chip-Technologie halb-quantitativ erfaßt. Ansätze des „Reverse Engineering“ adaptieren statistische Methoden, wie die Hauptkomponenten-, Cluster- oder Zeitreihenanalyse, um Korrelationen zwischen den Expressionsmustern verschiedener Gene aus den relativ großen Datensätzen abzuleiten und das unterliegende genetische Netzwerk zu erschließen (Eisen et al. 1998; Liang et al. 1998; Friedman et al. 2000; D’haeseleer et al. 2000; Wade und Hertz 2000). Daneben werden genregulatorische Netzwerke unter Verwendung verschiedener Modellierungs- und Simulationstechniken beschrieben, zu denen u.a. Graph-theoretische oder Regel-basierte Ansätze, Bayes- oder Boole’sche Netzwerke bzw. nichtlineare, partielle, stochastische und qualitative Differentialgleichungen zählen (McAdams und Arkin 1998; de Jong 2000; Smolen et al. 2000).
- Die Proteomanalyse („Proteomics“) beschreibt die Untersuchungen der in einem Organismus zu einem bestimmten Zeitpunkt vorhandenen Proteine und ihrer Wechselwirkungen, aus denen der Phänotyp der Zelle resultiert (Lee 2001). Die Proteinkonzentrationen und Interaktionen in den Zellen werden über Methoden, wie der 2D- bzw. 3D-Gelelektrophorese, der Massenspektrometrie, der 2-Hybrid-Technologie oder den Proteom-Chips, mit einer relativ geringen Genauigkeit quantifiziert (Alm und Arkin 2003). Die Daten sind jedoch hinreichend für die Ableitung von Protein-Protein-Interaktionsnetzwerken (Uetz et al. 2000; Ito et al. 2001) und die Analyse ihrer globalen Struktur (Jeong et al. 2001; Park et al. 2001). Eigenschaften der Topologie metabolischer Netzwerke, s. Abschnitt 2.2.2, können auch hier wiedergefunden werden (Rzhetsky und Gomez 2001).

In-Vitro^{2.5} Methoden zur Proteinanalyse beschäftigen sich mit der Isolierung und Charakterisierung einzelner Proteine. Dabei werden u.a. auch Daten zur enzymatischen Regulation, d.h. Substrate bzw. Effektoren von Enzymen, Reaktionskinetiken und enzymkinetische Parameter, erhoben, die in mechanistische Modelle für metabolische Netzwerke einfließen.

^{2.5} Untersuchungen im Reagenzglas.

- Die Aufklärung der drei-dimensionalen Struktur der Proteine über die relativ zeitaufwendige Methode der Röntgenstrukturanalyse wird als „Structural Genomics“ bezeichnet. Auf der Grundlage der Proteinstruktur kann durch Docking-Analysen (Westhead et al. 1997; Makino und Kuntz 1997; Apostolakis et al. 1998) die Bindungsaffinität der Proteine hinsichtlich verschiedener Liganden modelliert (Tokarski und Hopfinger 1997; Mandell et al. 2000), bzw. Substrat-Mimics durch Superposition (Lemmen und Lengauer 1997) gefunden werden. Auf dem Gebiet des „Molecular Modelling“ wird auch die Entstehung von Enzymkomplexen durch Protein-Protein-Wechselwirkungen untersucht. Über molekulardynamische Simulationsrechnungen können Erkenntnisse über den Funktionsmechanismus (Li et al. 1998; Oberhammer 1998) und die Umsetzungsgeschwindigkeiten des Enzym-Substrat-Komplexes gewonnen werden (Rarey et al. 1996; Eldridge et al. 1997).

Unter „Metabolomics“ (Tweeddale et al. 1998; Fiehn et al. 2000) versteht man die Untersuchungen der in einer Zelle vorhandenen niedermolekularen Bestandteile in unterschiedlichen Wachstumsphasen und unter variierenden experimentellen Bedingungen meist mittels der LC-MS-Analytik (LC-MS: Liquid-Chromatography Mass-Spectrometry). Hierbei kann es sich um Momentaufnahmen eines großen Metabolitspektrums (Raamsdonk et al. 2001) („metabolic fingerprinting“ bzw. „metabolic footprinting“) oder um die zeitlichen Verläufe einzelner Metabolite („metabolic profiling“) (Theobald et al. 1997b; Glassbrock et al. 2000; Buchholz et al. 2002) handeln. Die Modellierungsansätze auf diesem Gebiet werden in den folgenden Abschnitten näher dargestellt.

Die bisher auf das molekulare Detail konzentrierte Sichtweise der Biowissenschaften wird in letzter Zeit verstärkt durch Bemühungen abgelöst, das Verhalten und die Wechselwirkungen aller Elemente in einem funktionierenden biologischen System zu untersuchen (Agrawal 1999; Kitano 2000; Palsson 2000; Ideker et al. 2001a; Systeme des Lebens, Systembiologie (D) 2004; Genomes to Life (USA) 2004; Microbial Cell Project (USA) 2004; Complex Biological Systems Initiatives (USA) 2004; Alliance for Cellular Signaling (USA) 2004). Wolkenhauer (2001) definiert dieses Forschungsgebiet der Systembiologie als den „Einsatz systemtheoretischer^{2.6} Konzepte im Feld der Genomics^{2.7}“. Ein system-weites Verständnis genetischer oder metabolischer Reaktionsnetzwerke soll dabei durch den neuen Denkansatz erreicht werden, die Untersuchung von Beziehungen (Organisation und Struktur) bzw. Interaktionen (Dynamik und Verhalten) der gesamten Zelle in den Fokus der Betrachtungen zu rücken (Kitano 2002a; Kitano 2002b). Die integrierte Auswertung von Datensätzen aus dem Genom, Proteom und Metabolom (Hood 2001; Fiehn 2001; Ideker et al. 2001b), bzw. evtl. auch die Integration und der Austausch verschiedener existierender Modelle kann dazu hilfreich sein (Tomita et al. 1999; Hucka et al. 2003).

2.2.2 Modellierungsansätze zur Beschreibung metabolischer Netzwerke

Die Modellierung metabolischer Netzwerke wird seit mehreren Jahrzehnten, s. z.B. (Chance et al. 1960), mit unterschiedlicher Zielsetzung verfolgt, wobei sich die Modellierungsansätze von der Analyse der diskreten Struktur des Netzwerkes bis hin zur quantitativen Beschreibung des dynamischen Systemverhaltens erstrecken. In den folgenden Abschnitten ist eine grobe Klassifikation der existierenden Modelle gegeben. Die in dieser Arbeit verwendeten nichtlinearen Differentialgleichungssysteme zur Beschreibung des Metabolismus werden in Abschnitt 2.2.3 näher charakterisiert.

^{2.6}Der Begriff der Systemtheorie als formale Theorie der Beziehungen zwischen gekoppelten Systemen und ihrer Umgebung wurde 1940 vom Biologen Ludwig von Bertalanffy eingeführt. Sie ist eng verwandt mit der Kybernetik und charakterisiert eine neue Sichtweise, derzufolge ein System mehr auszeichnet als die Summe seiner Einzelbestandteile. Die Systemtheorie wurde sowohl in der Philosophie, z.B. von Bunge, Laslo, Bahm, als auch in der Mathematik, z.B. von Klir, weiterentwickelt und findet vielseitige Verwendung, wie z.B. in den Ingenieurwissenschaften oder in der Ökologie.

^{2.7}In diesem Zusammenhang stellt Genomics den Oberbegriff der vorherigen Definitionen der Genomics und der Functional Genomics dar.

Qualitative Modelle Die bekannteste Beschreibung des Stoffwechsels ist das zugrundeliegende biochemische Reaktionsnetzwerk, dessen einzelne Reaktionsschritte in der konventionellen chemischen Notation repräsentiert werden. Eine solche Grundstruktur ist zur Darstellung der Zellphysiologie geeignet, weil alle intrazellulären Vorgänge mit chemischen Stoffflüssen verbunden sind. Dies betrifft die Bildung und den Austausch:

- von Material, z.B. durch chemische Umsetzungen im Anabolismus oder Transportschritte zwischen Kompartimenten^{2,8};
- von Energie, z.B. durch die Produktion energiereicher Kosubstrate im Katabolismus;
- und von Information, wie die Synthese informationshaltiger Moleküle (DNS, RNS, etc.);

Graphen-theoretische Repräsentationen des Netzwerkes (Kohn und Lemieux 1991) stellen die Metabolite als Knoten dar, die über Kanten, d.h. die jeweiligen biochemischen Reaktionen oder genauer die betreffenden Enzym-Substrat-Übergangskomplexe, miteinander verbunden werden. Untersuchungen der Topologie des Netzwerkes ergaben, daß die Wahrscheinlichkeitsverteilung für die Konnektivität der Metabolite einem Potenzgesetz folgt, d.h. das Netzwerk skalenfrei ist (Barabasi und Albert 1999). Eine weitere, universelle Eigenschaft der Netzwerke ist ihr „Small-world“-Charakter. Darunter versteht man, daß zwei beliebige Knoten auf relativ kurzen Pfaden über existierende Kanten miteinander verbunden werden können (Watts und Strogatz 1998; Fell und Wagner 2000). Die durchschnittliche Länge des Pfades, der Netzwerkdiаметer, ist bei Betrachtung verschiedener Organismen sehr ähnlich (Jeong et al. 2000). Die modulare Struktur des Netzwerkes ist Gegenstand weiterer Untersuchungen, die sich vorwiegend mit der Dekomposition des Graphen beschäftigen (Hartwell et al. 1999; Schuster et al. 2002; Gagneur et al. 2003).

Bipartite Graphen oder Digraphen bestehen aus Stoff- und Reaktionsknoten, auf deren Grundlage einige Netzwerkeigenschaften, wie Erhaltungsgrößen, Stöchiometrie und Reaktionspfade, untersucht werden (Kohn und Lemieux 1991; Wiechert 1994; Asratian et al. 1998). Die Struktur kann durch die Hinzunahme regulatorischer Einflüsse von Effektoren erweitert werden, s. Abb. 2–3. Auch einfache oder hybride Petrinetze werden zur Graph-orientierten Beschreibung der Netzwerke verwendet (Reddy et al. 1996; Küffner et al. 2000). Im einfachen Petrinetz sind zwischen den zu den Metabolitkonzentrationen analogen Stellen des Petrinetzes Transitionen von Marken (Repräsentation für die einzelnen Moleküle) über die verbindenden Kanten (Reaktionen) möglich, so daß Stoffflüsse auf eine diskrete Weise dargestellt werden können.

Ereignisdiskrete Beschreibungsansätze, z.B. stochastische Simulationen (Boiteux et al. 1975; Jones et al. 2002), automaten-theoretische Konzepte (Thomas 1991; Garcia-Olivares et al. 2000; Siehs et al. 2002; Yang 2003) und die diskret-ereignisorientierte Modellierung (Degenring et al. 2003; Xia und Wise 2003; Degenring et al. 2004b), werden zur semi-quantitativen bis quantitativen Beschreibung des dynamischen Verhaltens des Netzwerkes genutzt. Bei diesen Ansätzen wird das System in Bestandteile, wie Metabolite, Enzyme, etc., mit verschiedenen Zuständen aufgeteilt, zwischen denen diskrete Übergänge möglich sind. In einigen Fällen werden die Enzyme sogar als Agenten angesehen, die bewußte Entscheidungen treffen können (Jonker et al. 2002).

Logische Modelle, d.h. Boole'sche, Fuzzy-logische oder Regel-basierte Ansätze, werden ebenfalls für die Beschreibung der Dynamik der Enzymreaktionen eingesetzt (Arkin und Ross 1994; Hjemfelt und Ross 1995; Hauri et al. 1997), z.B. wenn Daten für mechanistische Modelle fehlen.

^{2,8}Räumliche Strukturen im Zellinneren eukaryotischer Zellen, d.h. Organellen wie Mitochondrien oder Chloroplasten, die mit Membranen vom Cytoplasma abgetrennt sind;

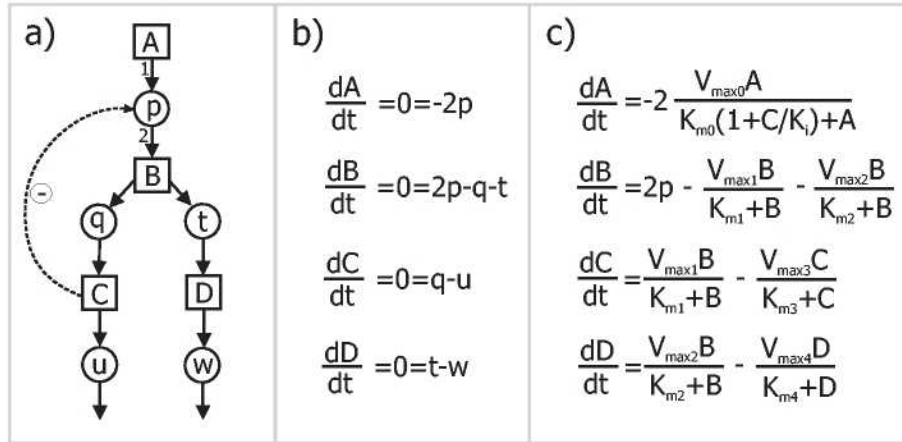


Abb. 2-3 Verschiedene Modell-basierte Repräsentationen des metabolischen Netzwerkes; a) bipartiter Graph mit Wechselwirkungsrelationen; b) stöchiometrische Stoffbilanzen; c) nichtlineares Differentialgleichungsmodell

Stöchiometrische Modelle Unter der Annahme, daß nur kleine Zeiträume betrachtet werden und die Umgebungsbedingungen sich langsam ändern, ist es möglich, die Konzentrationen aller Metabolite innerhalb des Netzwerkes als nahezu konstant zu erachten. Somit können die Änderungsraten der einzelnen Metabolitkonzentrationen, s. Abb. 2-3 b, auf Null gesetzt und Bilanzgleichungen für die Geschwindigkeiten der Reaktionen, an denen die jeweiligen Metabolite beteiligt sind, aufgestellt werden, s. (Heinrich und Schuster 1996; Fell 2003). Die Notation der Bilanzgleichungen in Matrix-Form lautet:

$$\mathbf{0} = \mathbf{N} \cdot \mathbf{v}, \quad (2-1)$$

mit $\mathbf{v}$ dem Vektor der Reaktionsgeschwindigkeiten und $\mathbf{N}$ der Matrix der stöchiometrischen Koeffizienten. Sie sind für das Beispiel in Abb. 2-3 folgendermaßen definiert (Stephanopoulos et al. 1998):

$$\mathbf{N} = \begin{array}{ccccc} \begin{array}{c} \text{Reaktionen} \\ \leftarrow r_{AB} \quad r_{BC} \quad r_{BD} \quad r_C \quad r_D \end{array} & \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 2 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix} & \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ \text{D} \end{array} \\ \text{Metabolite} & & \end{array} \quad (2-2)$$

$$\mathbf{v} = (p \quad q \quad t \quad u \quad w)^T. \quad (2-3)$$

Die Spalten der stöchiometrischen Matrix sind einer bestimmten Reaktion zugeordnet und beschreiben, welche Metabolite an der Reaktion beteiligt sind und in welcher Menge sie entstehen bzw. verbraucht werden. Der Flußvektor $\mathbf{v}$ beinhaltet die Gleichungen für die Reaktionsgeschwindigkeiten in der Reihenfolge, in der die Spalten der Matrix sortiert sind.

Auf Basis der Bilanzgleichungen können mit Methoden der Stoffflußanalyse intrazelluläre Stoffflüsse im stationären Zustand aus extrazellulär bestimmten Flüssen, wie z.B. Aufnahmeraten von Substraten, berechnet werden (Varma und Palsson 1994; Nissen et al. 1997; Vallino und Stephanopoulos 2000). Oft ist das Gleichungssystem unterbestimmt, d.h. die Zahl der gemessenen Flüsse ist kleiner als der

Freiheitsgrad des Systems, und es gibt eine unendliche Menge an Lösungen für die Flüsse im Netzwerk. In diesem Fall wird die Lineare Optimierung eingesetzt, um über die Minimierung oder Maximierung einer Zielfunktion optimale Flußverteilungen bzw. sogar eine einzigartige Lösung zu ermitteln (Varma et al. 1993; Pramanik und Keasling 1997; Edwards und Palsson 2000a; Edwards und Palsson 2000b). Informationen aus Kohlenstoff-Markierungsexperimenten, bei denen ^{13}C oder ^{14}C -markierte Substrate verwendet und die Isotopenmuster der intrazellulären Metabolite über die NMR^{2.9}- (Schmidt et al. 1997; Szyperski 1998) bzw. GC-MS^{2.10}- (Christensen und Nielsen 1999) Analytik bestimmt werden, können ebenfalls zur Einschränkung des Lösungsraums benutzt werden (Wiechert und de Graaf 1997; Wiechert et al. 1997; Christensen und Nielsen 2000). Basierend auf den Ergebnissen der Stoffflußanalyse ist es möglich, Beziehungen zwischen dem Phänotyp und dem Genotyp eines Organismus aufzustellen (Schilling et al. 1999a; Price et al. 2002b).

Anstatt spezielle Lösungen für die Flußverteilungen zu berechnen, gelingt es über eine Konvexanalyse zur Ermittlung der elementaren Flußmoden oder der Bestimmung eines Basissatzes von Vektoren, die den Nullraum der Matrix aufspannen, Einsichten über die Struktur des Reaktionsnetzwerkes zu erhalten (Schilling et al. 1999b; Schuster et al. 1999; Schuster et al. 2000; Stelling et al. 2002).

Mechanistische Modelle Bei sich rasch ändernden Umgebungsbedingungen oder Beobachtungen über einen hinreichend langen Zeitraum sind die Vereinfachungen der stöchiometrischen Modelle nicht länger zulässig, so daß weder die Metabolitkonzentrationen noch die Reaktionsgeschwindigkeiten mehr als konstant erachtet werden dürfen. Dies führt zu einem System nichtlinearer, gewöhnlicher Differentialgleichungen, das in Matrix-Notation wie folgt formuliert werden kann:

$$\dot{\mathbf{c}} = \mathbf{N} \cdot \mathbf{v}(\mathbf{p}, \mathbf{E}, \mathbf{c}_{\text{ex}}, \mathbf{c}). \quad (2-4)$$

Die Reaktionsgeschwindigkeiten sind von dem Vektor der intrazellulären und der extrazellulären Konzentrationen $\mathbf{c}$ bzw. $\mathbf{c}_{\text{ex}}$ dem Vektor der Enzymkonzentrationen $\mathbf{E}$ und dem Parametervektor $\mathbf{p}$ abhängige Funktionen. Die Abhängigkeit der Reaktionsgeschwindigkeit von den einzelnen Faktoren wird häufig über enzymkinetische Terme (Segel 1993; Segel 1999; Bisswanger 2000; Cornish-Bowden 2004), wie z.B. der Michaelis-Menten-Gleichung für die irreversible Umsetzung eines Substrates zu einem Produkt, beschrieben:

$$v = \frac{V_{\text{max}} \cdot C}{K_m + C}. \quad (2-5)$$

Die Reaktionsgeschwindigkeit v ist eine Funktion der Metabolitkonzentration^{2.11}, C , und zweier für das Enzym spezifischer Parameter V_{max} und K_m . Für hohe Metabolitkonzentrationen, C , zeigt die Reaktionsgeschwindigkeit ein Sättigungsverhalten, wobei C mit der maximalen Reaktionsgeschwindigkeit, $V_{\text{max}} = k_{\text{cat}} \cdot E$ umgesetzt wird^{2.12}. Bei geringen Konzentrationen des Substrates ist die Geschwindigkeit proportional zum Quotienten aus V_{max} und K_m , der Michaelis-Menten-Konstante. Enzymkinetiken, die darüber hinaus die Reversibilität von Reaktionen und die Einflüsse von Effektoren berücksichtigen, können eine hohe Anzahl von Parametern (10-20) beinhalten, s. Anhang D.

^{2.9}NMR: Nuclear Magnetic Resonance Spectrometry.

^{2.10}GC-MS: Gas Chromatography Mass Spectrometry.

^{2.11}Um die mathematischen Formeln übersichtlicher zu gestalten, wird in dieser Arbeit auf die gängige chemische Notation der Konzentrationen mit eckigen Klammern verzichtet.

^{2.12} k_{cat} beschreibt die für das Enzym charakteristische Frequenz für die Umsetzung eines Metaboliten, E die Enzymkonzentration. E kann häufig nicht exakt bestimmt werden, so daß sie meist als implizite Größe, der Aktivität, über den Parameter V_{max} in die kinetischen Gleichungen einfließt.

Das Differentialgleichungssystem für das in Abb. 2–3 gegebene Beispiel lautet nach Einsetzen der Enzymkinetiken und der stöchiometrischen Matrix (Gleichung (2–2)) in (2–4):

$$\begin{pmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \\ \frac{dD}{dt} \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 \\ 2 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \frac{V_{\max,0}A}{K_{m,0}(1 + C/K_{i,c}) + A} \\ \frac{V_{\max,1}B}{K_{m,1} + B} \\ \frac{V_{\max,2}B}{K_{m,2} + B} \\ \frac{V_{\max,3}C}{K_{m,3} + C} \\ \frac{V_{\max,4}D}{K_{m,4} + D} \end{pmatrix} \quad (2-6)$$

2.2.3 Mechanistische Modelle zur Untersuchung der Stoffwechseldynamiken

Instationäre mechanistische Modelle können erstellt werden, wenn detaillierte Informationen über die Kinetik bestimmter zellulärer Prozesse vorhanden sind. Sie werden wie in dieser Arbeit vor allem eingesetzt, um das globale dynamische Verhalten des Stoffwechsels zu untersuchen. In den folgenden Abschnitten ist eine Übersicht über verschiedene Modelle und Methoden zur Modellanalyse bzw. Vereinfachung gegeben.

Modelle Neben den Modellierungsansätzen, die die zeitliche Veränderung der molekularen Komposition, Gestalt und Größe einer einzelnen Zelle mit einem nichtlinearen Differentialgleichungssystem (DGL-System) beschreiben (Domach et al. 1984; Laffend und Shuler 1994), gibt es zahlreiche Modelle, die für die Voraussage des über viele Zellen gemittelten Verlaufs der Metabolitkonzentrationen erstellt wurden. Unstrukturierte Modelle behandeln die Zelle als „Black Box“, wobei intrazelluläre Vorgänge vernachlässigt und nur Flüsse in und aus der Zelle betrachtet werden, z.B. in (Sonnleitner und Käppeli 1986). Strukturierte Modelle können entstehen, wenn verschiedene zelluläre Prozesse zu funktionalen Einheiten (Modulen) gruppiert werden, die über Schnittstellen miteinander verbunden sind (Kremling et al. 2000; Kremling und Gilles 2001; Kremling et al. 2001)^{2,13}. Die im Folgenden vorgestellten Modelle sind räumlich strukturiert, d.h. es wird zwischen den Konzentrationen eines Metaboliten in verschiedenen Kompartimenten differenziert. Unter ihnen setzen einige Modelle *in-vitro* bestimmte Enzymkinetiken mit unveränderten Parameterwerten ein, um das Verhalten der lebenden Zellen zu simulieren und die Simulationen mit den experimentellen Ergebnissen zu vergleichen (Bakker et al. 1997; Teusink et al. 2000). Häufig werden dabei größere Diskrepanzen festgestellt (Wright und Kelly 1981; Wright und Albe 1994), weil u.a.:

- die *in-vivo*^{2,14} Enzymkonzentrationen wesentlich höher sind als bei *in-vitro* Experimenten, was z.B. vermehrte Enzym-Enzym Interaktionen ermöglicht. Dies kann aufgrund von Volumen-Ausschluss-effekten zu einem nicht-idealen Verhalten führen, d.h. die Enzymaktivität steigt überproportional zur Enzymkonzentration an, und die Dissoziationsgeschwindigkeit der Enzymkomplexe sinkt (Rohwer et al. 1998). Mit diesem Phänomen des „Macromolecular Crowding“ eng verwandt ist das „Metabolite Channeling“ (Agius und Sheratt 1997; Huang et al. 2001), bei dem die Intermediate der Reaktionen direkt von dem aktiven Zentrum eines Enzyms auf das nächste übertragen werden,

^{2,13}In die funktionalen Einheiten wurden auch Vorgänge der Transkription und Translation der entsprechenden Gene integriert, wobei eine hierarchische Organisation der Signaltransduktion festgestellt wurde.

^{2,14}Im lebendem System.

wobei nur geringe Mengen zwischenzeitlich in das Cytoplasma freigesetzt werden;

- die *in-vivo* Konzentration der Substrate sehr viel geringer ist als die *in-vitro* gewählten Konzentrationen;
- unter *in-vivo* Bedingungen eine große Zahl zusätzlicher Metabolite vorhanden ist, die die Aktivität der Enzyme beeinflussen können;
- physikalische Größen wie der pH-Wert, die Temperatur oder Salzkonzentrationen bei *in-vitro* Experimenten häufig anders sind als unter physiologischen Bedingungen;

Um daher das Verhalten der Zelle besser beschreiben zu können, werden zwar noch die Gleichungen der *in-vitro* bestimmten Kinetiken in den Modellen verwendet, aber die Parameterwerte der Kinetiken optimiert, indem die simulierten Trajektorien an experimentelle Daten aus *in-vivo* Messungen angepaßt werden. Auf diese Weise resultierten z.B. Modelle für die Glykolyse (Galazzo und Bailey 1990; Cortassa und Aon 1994; Hynne et al. 2001) und den Penicillin-Biosyntheseweg in *Penicillium chrysogenum* (Nielsen und Jørgensen 1996; Pissara et al. 1996). Die Modelle werden teilweise für biotechnologische Anwendungszwecke eingesetzt, wie im letzten Fall, daneben aber auch, um das biologische Verhalten, wie z.B. die Oszillationsfähigkeit des Systems (Sel'kov 1968; Goldbeter und Lefever 1972; Richter et al. 1975; Termonia und Ross 1981; delaFuente et al. 1996; Nielsen et al. 1998; Wolf und Heinrich 2000; Dano et al. 2001; Wolf et al. 2001), zu charakterisieren.

Zur Auswertung einer speziellen Art von *in-vivo* Messungen, auf denen auch diese Arbeit basiert, s. Abschnitt 3.2, wurden Modelle für die Glykolyse (Theobald et al. 1997a), den Pentose-Phosphat-Weg (Vaseghi et al. 1999) in *Saccharomyces cerevisiae*, das Phosphotransferase-System (pts) (Sauter et al. 2002) und den Zentralstoffwechsel in *Escherichia coli* (Chassagnole et al. 2002) entwickelt. Diese Modelle mit einer Dimension von bis zu 100 DGL und mehr als 80 Parametern sind auf die Identifikation metabolischer Regulatoren und der *in-vivo* Parameterwerte fokussiert.

Analyse der Modelle Die Analysen der DGL-Modelle beschäftigen sich meist mit der Untersuchung eines Fließgleichgewichtszustandes, den diese bei einem vorgegebenen Parametersatz (z.B. aus einer vorhergehenden Optimierung) erreichen können:

$$\mathbf{0} = \mathbf{N} \cdot \mathbf{v}(\mathbf{p}, \mathbf{E}, \mathbf{c}_{\text{ex}}, \mathbf{c}). \quad (2-7)$$

Ausgehend von dieser Darstellung werden stationäre Zustände berechnet (Hofmeyr 1986), sowie Sensitivitäts- und Bifurkationsanalysen^{2,15} durchgeführt (Segel 1988; Hynne et al. 2001; Davidson et al. 2002).

Die metabolische Kontrollanalyse, „Metabolic Control Analysis“ (MCA), ist eine spezielle Ausprägung der Sensitivitätsanalyse und wurde basierend auf zwei unabhängigen Arbeiten (Kacser und Burns 1973; Heinrich und Rapoport 1974a) in (Burns et al. 1985) zu einem theoretischen Gerüst zusammengeführt. Sie untersucht für den Fließgleichgewichtszustand, für den Gleichung (2-7) gilt, wie sich die Flüsse bzw. Konzentrationen in Abhängigkeit von Störungen des Systems (z.B. durch Änderungen der Enzymaktivitäten) ändern und diese Eigenschaften mit einzelnen Komponenten des Systems (bestimmten Enzymen) oder strukturellen Eigenschaften des Netzwerkes in Verbindung gebracht werden können. Verschiedene Koeffizienten wurden definiert, um die Auswirkungen infolge der Störungen zu quantifizieren. Sie sind dimensionslos, weil sie auf den Referenzzustand des Fließgleichgewichtes mit den

^{2,15}Der Begriff Bifurkation oder Verzweigung bezeichnet den Übergang von einem Systemzustand in einen qualitativ anderen als Folge einer i.A. stetigen Änderung eines oder mehrerer Kontrollparameter $\mu \in \mathbb{R}$. Ein Punkt μ_c im Parameterraum, bei dem qualitativ neue Bewegungstypen auftreten, heißt Bifurkationspunkt. Konkret bedeutet das, daß das (Stabilitäts-)Verhalten eines Fixpunktes oder anderer Attraktoren (beide stellen Lösungen, $\mathbf{c}(t)=\mathbf{c}$, von (2-7) dar) in Abhängigkeit von Veränderungen des Parametervektors $\mathbf{p}$ untersucht wird.

Enzymkonzentrationen, E_z^0 , den Konzentrationen, c_j^0 , und den Flüssen durch lineare, unverzweigte Abschnitte des Reaktionsnetzwerkes, v_h^0 , bezogen werden. Die globalen Fluß- bzw. Konzentrationskontrollkoeffizienten, $C_{hz}^{v_0}$ bzw. $C_{jz}^{c_0}$, geben wieder, wie ein Fluß, v_h , oder eine Konzentration, c_j , durch die Änderung einer Enzymkonzentration, E_z , beeinflusst wird, s. auch (Fell 2003):

$$C_{hz}^{v_0} = \frac{E_z^0}{v_h^0} \left(\frac{dv_h}{dE_z} \right)^0, \quad C_{jz}^{c_0} = \frac{E_z^0}{c_j^0} \left(\frac{dc_j}{dE_z} \right)^0. \quad (2-8)$$

Analog zu dieser Darstellung quantifizieren Response-Koeffizienten die Beeinflussung der Konzentrationen bzw. Flüsse infolge einer kleinen Änderung eines Parameters, wie z.B. einer extrazellulären Konzentration. Die Elastizitäten, ε^0 sind lokale Koeffizienten und beschreiben, wie sich eine bestimmte Reaktionsgeschwindigkeit, v_h , bei Variation einer Konzentration, c_j , einer Enzymkonzentration, E_z , oder eines Parameters, wie z.B. $c_{ex,k}$ oder p_g , ändert, wobei im Gegensatz zu den globalen Koeffizienten alle anderen Größen konstant gehalten werden:

$$\varepsilon_{hj}^{c_0} = \frac{c_j^0}{v_h^0} \left(\frac{\partial v_h}{\partial c_j} \right), \quad \varepsilon_{hz}^{E_0} = \frac{E_z^0}{v_h^0} \left(\frac{\partial v_h}{\partial E_z} \right), \quad \varepsilon_{hk}^{c_{ex,0}} = \frac{c_{ex,k}^0}{v_h^0} \left(\frac{\partial v_h}{\partial c_{ex,k}} \right). \quad (2-9)$$

Zwischen den Koeffizienten existieren für jedes Netzwerk gültige Beziehungen (Reder 1988), wie das Summationstheorem für die Addition aller Fluß- bzw. Konzentrationskontrollkoeffizienten zu 1 bzw. 0 und das Konnektivitätstheorem, das Elastizitäten und Kontrollkoeffizienten miteinander verknüpft.

Gemäß der Definitionen für die Koeffizienten in den Gleichungen (2-8) und (2-9) ist die klassische MCA eine lineare Störungstheorie^{2,16} und gilt damit nur nahe des Fließgleichgewichtszustandes und für kleine Parameteränderungen. Ungeachtet dieser Einschränkungen ist die Methode ein häufig verwendetes Werkzeug für das „Metabolic Engineering“, weil z.B. der Einfluß bestimmter Reaktionen auf die Flußverteilung berechnet und die Effekte von Enzymaktivitäten auf intrazelluläre Metabolitkonzentrationen beschrieben werden können (Galazzo und Bailey 1990; Cortassa und Aon 1994; Schlosser et al. 1994; Nielsen und Jørgensen 1995; Pissara et al. 1996). Erweiterungen der MCA formulieren zeitabhängige Kontrollkoeffizienten für nicht-stationäre Zustände (Kholodenko et al. 1997; Mauch et al. 1997; Hatzimanikatis 1999; Ingalls und Sauro 2002).

Die „Biochemical Systems Theory“ (BST) (Savageau 1969a; Savageau 1969b; Savageau 1970) ist ein mit der MCA eng verwandter Ansatz (Savageau et al. 1987a; Savageau et al. 1987b), der davon ausgeht, daß die Reaktionsgeschwindigkeiten über Potenzgesetze repräsentiert werden können:

$$v_h = \kappa_h \prod_j y_j^{\zeta_{hj}} \quad (2-10)$$

y_j bezeichnet die Zustandsvariablen des Systems (Metabolitkonzentrationen, Enzymaktivitäten, Effektoren, etc.), κ_h die apparenten Geschwindigkeitskonstanten und ζ_{hj} die apparenten Ordnungen der Kinetiken. Entsprechend der Theorie werden alle Reaktionen, die einen Metaboliten c_j generieren bzw. verbrauchen, zu zwei Netto-Reaktionen mit der Geschwindigkeit v_h bzw. v_{-h} zusammengefaßt. Würden logarithmische Variablen ($\ln(c_j)$) in (2-10) eingesetzt, so entstünde ein System linearer Gleichungen, d.h. der BST-Ansatz linearisiert ähnlich zur MCA die nichtlinearen Enzymkinetiken auf einer logarithmischen Skala. Aufgrund der Massenbilanzen der Metabolite kann dann ein System von DGL, das synergistische oder S-System (Savageau 1985; Voit 1991), aufgeschrieben werden, das in Analogie zur MCA hauptsächlich verwendet wird (Hatzimanikatis und Lee 1999), um Flüsse und die Flußverteilung

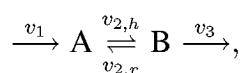
^{2,16}Die Taylor-Reihenentwicklung für die Ableitungsfunktion der nichtlinearen Enzymkinetiken wird nach dem ersten Glied abgebrochen, d.h. die Funktion wird linearisiert, s. Abschnitt 6.2.2 für eine genauere Darstellung.

kontrollierende Einflüsse zu quantifizieren. Im Gegensatz zur klassischen MCA können über die BST auch transiente Zustände untersucht werden, die Interpretation der Parameterwerte im biochemischen Kontext ist jedoch bei diesem Ansatz erschwert.

Vereinfachung der Modelle Aus der Beschreibung des Prozesses zur Modellerstellung, s. oben, wird deutlich, daß sich die Dimension der Modelle, d.h. die Zahl ihrer DGL und Parameter, in Abhängigkeit von dem dargestellten Reaktionsnetzwerk rasch vergrößern kann. Dadurch ergeben sich unter Umständen technische Probleme, die Vereinfachungen des Modells sinnvoll erscheinen lassen. Unter methodischen Gesichtspunkten sind vereinfachende Annahmen unerlässlich, weil ein Modell immer zur Beantwortung einer bestimmten Fragestellung (z.B. zur Auswertung eines konkreten Datensatzes) entwickelt wird, d.h. nur innerhalb eines bestimmten experimentellen Rahmens (Zeigler et al. 2000) gültig ist. Einige gängige Methoden zur Reduktion der Dimension des DGL-Systems sind im Folgenden dargestellt.

Reaktionsnetzwerke können vereinfacht werden, indem Reaktanden, Reaktionen oder beides zusammengefaßt werden (Zeigler 1988), s. Anhang E. Dabei werden Bilanzgleichungen genutzt, um die Konzentrationen und Flüsse der ursprünglichen Reaktanden bzw. Reaktionen näherungsweise mit denen der kombinierten auszugleichen. Ein weiteres Hilfsmittel zur Vereinfachung der Reaktionsmechanismen stellt die Analyse der Zeitskalen, innerhalb derer verschiedene metabolische Reaktionen ablaufen, dar (Delgado und Liao 1995). Von denen sich im Rahmen des Beobachtungszeitraumes abspielenden Prozessen, die demzufolge mit DGL modelliert werden sollten, unterscheiden sich folgende Prozesse (Heinrich und Schuster 1996; Vanrolleghem und Heijnen 1998):

- Spezies im quasi-stationären Zustand: Metabolite mit einer hohen Umsetzungsrate, deren Akkumulation vernachlässigt werden kann, weil die Netto-Bildungs- der Netto-Verbrauchsrate gleichkommt. Die Differentialgleichungen für solche Metabolite dürfen aus dem Modell eliminiert werden (Bodenstein 1913).
- Reaktionen in partiellen Gleichgewichten: Reaktionen, die so schnell ablaufen, daß eine sofortige Einstellung des Gleichgewichtes bzgl. ihrer Reaktanden angenommen werden kann ($v = \infty$). Damit ist es möglich, die Geschwindigkeiten der schnellen Reaktionen zu eliminieren, indem die Gleichgewichtsbeziehungen für die Berechnung der Metabolite im Pseudo-Gleichgewicht genutzt werden. Verläuft z.B. in einem hypothetischem Reaktionsmechanismus:



die Reaktion von A nach B sehr schnell, können die Metabolite A und B zu einer Summengröße, C, kombiniert werden:

$$\begin{aligned} \frac{dA}{dt} &= v_1 - v_{2,h} + v_{2,r}; & \frac{dB}{dt} &= v_{2,h} - v_{2,r} - v_3; & \text{mit } v_{2,h}, v_{2,r} &\ll v_1, v_3 \\ \Rightarrow \frac{d(A+B)}{dt} &= \frac{dC}{dt} = v_1 - v_3, \end{aligned}$$

und die ursprünglichen Konzentrationen über die Beziehungen $C = A + B$ und $K_{eq} = B/A$ zu jeder Zeit aus C berechnet werden (Heinrich und Schuster 1996).

- Sehr langsame Reaktionen: solche Reaktionen haben sehr große Zeitkonstanten im Vergleich zum Zeitfenster der Beobachtungen ($v = 0$), so daß sie vernachlässigt werden dürfen. Dies kann zum Auftreten von Erhaltungsgrößen führen. Z.B. ist die Summenkonzentration der Adenosin-Nukleotide über die Zeit hinweg konstant, $ATP + ADP + AMP = \text{konst.}$, wenn die Synthese und der Abbau der Nukleotide als langsam im Vergleich zu der gegenseitigen Umwandlung angesehen

wird. Aufgrund der Erhaltungsgleichung wird weiterhin die DGL einer der bilanzierten Komponenten überflüssig.

Über die zeitskalenbasierte Reduktion gelang es, das in (Theobald et al. 1997a) beschriebene Modell von 97 auf 27 DGL und von 83 auf 15 Reaktionskinetiken zu kürzen (Visser et al. 2000). Bei einem solchen Reduktionsprozeß ist allerdings zu berücksichtigen, daß Reaktionen, deren Effektoren aufgrund von vereinfachenden Annahmen eliminiert wurden, evtl. nicht mehr adäquat beschrieben werden können.

Neben den oben beschriebenen Methoden zur Reduktion des Reaktionsnetzwerkes gibt es verschiedene Ansätze, um die Zahl der Parameter in den enzymkinetischen Ausdrücken zu verringern. Geschwindigkeitsgesetze mit ganz- oder gebrochen-rationalen Ordnungen bzgl. der einzelnen Reaktanden („mass action kinetics“ (Smith et al. 1991)) stellen die einfachste Beschreibung der Enzymkinetiken dar. In diesem Fall wird das Geschwindigkeitsgesetz aus der Wahrscheinlichkeit, mit der sich verschiedene Reaktanden gemäß der Stöchiometrie der Reaktion treffen, und einer Geschwindigkeitskonstanten, die die Frequenz der Reaktion in Abhängigkeit von Temperatur, Druck, etc. angibt, zusammengesetzt:

$$1 A + 2 B \longrightarrow C \Rightarrow -\frac{1}{2} \frac{dA}{dt} = -\frac{dB}{dt} = v = k \cdot A \cdot B^2 \quad (2-11)$$

Eine Formulierung für die Enzymkinetiken über Potenzgesetze (s. BST) erweitert Gleichung (2-11) dadurch, daß die Ordnungen der Geschwindigkeiten auch die Konzentrationen anderer Reaktanden beinhalten können, um z.B. allosterische Effektoren zu berücksichtigen. In (Visser et al. 2000) werden beide Ansätze zu einer Gleichung („Tendency kinetics“) kombiniert, wodurch das in (Theobald et al. 1997a) beschriebene Modell von 84 auf 35 Parameter reduziert werden konnte. Gleichung (2-11) könnte dabei z.B. zu:

$$v = k \cdot A \cdot B^2 \cdot C^{-\zeta_C}, \quad (2-12)$$

erweitert werden, um einen evtl. inhibierenden Effekt von C zu berücksichtigen.

Linlog-Kinetiken (Westerhoff und van Dam 1987; Nielsen 1997) sind Ausdrücke, wie:

$$v = E(a + b \ln A + c \ln B + c \ln C), \quad (2-13)$$

in denen die Geschwindigkeit als proportional zur Enzymaktivität E und zur Summe nichtlinearer logarithmischer Terme für die Metabolitkonzentrationen ausgedrückt wird. Auf diese Weise können Enzymaktivitäten über einfache mathematische Transformationen in Relation zu Metabolitkonzentrationen und Flüssen im Fließgleichgewicht gebracht werden, was Ergebnisse für das Metabolic Engineering liefert (Hatzimanikatis et al. 1998; Visser und Heijnen 2003).

Neben den vorgestellten Formen nicht-mechanistischer Enzymkinetiken gibt es Geschwindigkeitsgesetze, die auf biochemisch bekannten Ausdrücken beruhen und an anderer Stelle vereinfacht werden. Multiplikative Gleichungen werden eingesetzt, um die Parameterzahl der Kinetiken für Mehrsubstrat-Reaktionen zu reduzieren, indem einfache Kinetiken von Monosubstrat-Reaktionen für verschiedene Reaktanden multipliziert werden. Die folgende Gleichung (Doppelsubstrat-Kinetik) kann die Ordered-Bi-Bi-Kinetik für die irreversible Reaktion zweier Substrate, s. Gleichung (D-12) auf S. 171 (Bisswanger 2000), näherungsweise ersetzen, wobei die Anzahl der Parameter reduziert wird:

$$v = \frac{V_{\max} \cdot A \cdot B}{A \cdot B + K_{mA} \cdot B + K_{mB} \cdot A + K_{iA} \cdot K_{mB}} \quad (\text{irrev. Ordered-Bi-Bi-Kinetik}),$$

$$\approx v = V_{\max} \cdot \frac{A}{K_{mA} + A} \cdot \frac{B}{K_{mB} + B} \quad (\text{Doppelsubstrat-Kinetik}).$$

Kybernetische Variablen werden mit dem Ziel in ein kinetisches Modell eingeführt, mechanistische Details durch Objektivfunktionen zu ersetzen, wobei angenommen wird, daß der Metabolismus auf gewisse Ziele hin arbeitet, wie das Erreichen der Homeostasis^{2,17} oder einer optimalen Wachstumsrate (Varner und Ramkrishna 1999a; Varner und Ramkrishna 1999b; Giuseppin und van Riel 2000). Der Ansatz wurde z.B. genutzt, um die Flußveränderungen der Threonin-Biosynthese genetisch modifizierter *Corynebacterium lactofermentum*-Organismen vorherzusagen (Varner und Ramkrishna 1999c). Der MILP-Ansatz (Mixed Integer Linear Programming Ansatz) ersetzt kinetische Ausdrücke durch diskrete Variablen, die die An- oder Abwesenheit von regulatorischen Schleifen repräsentieren (Hatzimanikatis et al. 1998). Fuzzy-Logik-basierte Modellierungsansätze (Lee et al. 1999; Babuška et al. 1999) benutzen algebraische Gleichungen, wie die Michaelis-Menten-Kinetik, für die Reaktionskinetiken und beschreiben z.B. die Variation der Reaktionsgeschwindigkeit in Abhängigkeit von unterschiedlichen Inhibitorkonzentrationen, I , über Fuzzy-Logik-Regeln:

$$v = V_{\max} \cdot \frac{A}{K_{mA} + A} \quad V_{\max} = \begin{cases} V_{\max,1} & \text{für } I < K_i \\ V_{\max,2} & \text{für } I \geq K_i \end{cases}, \quad (2-14)$$

(Michaelis-Menten-Kinetik)

(Fuzzy-Logik-Regel für den $V_{\max}$ -Wert der Kinetik)

Die Menge an experimentellen Daten, die für einen solchen Ansatz benötigt wird, steht jedoch meist in keinem guten Verhältnis zu den Erkenntnissen, die aus solchen Modellen gewonnen werden können.

2.3 Verifizieren, Validieren und Testen von Modellen

2.3.1 Allgemeines

Modelle werden entwickelt, um bestimmte Fragestellungen zu beantworten und auf der Basis von Modellexperimenten Entscheidungen zu treffen. Gleichzeitig ist der Vorgang der Modellentwicklung komplex, d.h. er besteht aus zahlreichen, eng verzahnten Einzelschritten, von denen jeder einzelne fehlerbehaftet sein kann (Page 1991). Um glaubwürdige Lösungen zu erhalten, die zur Ableitung von Folgerungen bzw. zur Entscheidungsfindung genutzt werden können, ist es also in Anbetracht der Komplexität des Modellierungsprozesses wichtig zu prüfen, ob das zugrundeliegende Modell „richtig“ ist (Balci 1998). Da ein Modell immer die Abstraktion eines Systems darstellt, wird es nie ein einziges, wahres oder vollkommen richtiges Modell und demzufolge auch keinen exakten Test geben, mit dem sich die Gültigkeit eines Modells endgültig beweisen läßt. Bei der Prüfung von Modellen kann es vielmehr nur darum gehen, größtmögliche Transparenz für die Beurteilung von Modellen zu schaffen und alle Möglichkeiten auszuschöpfen, inkorrekte Modelle zu falsifizieren. Hierfür werden Techniken zur Verifikation, Validierung und zum Testen eines Modells (VV&T) eingesetzt (Balci 1998).

Bei der *Verifikation* eines Modells wird untersucht, ob ein Modell korrekt erstellt wurde. Die Transformation eines Modells von einer Form in eine andere, z.B. die Umwandlung eines mathematischen Modells in ein ausführbares Computerprogramm (Implementation), sollte dabei mit hinreichender Genauigkeit erfolgt sein.

Die *Validierung* eines Modells beschäftigt sich damit, ob ein im Rahmen seines Anwendungsbereiches gültiges Modell erstellt wurde (Fishman und Kiviat 1968; Schlesinger 1979). Ein validiertes Modell verhält sich konsistent zu dem untersuchten System, das simulierte Verhalten ist z.B. vergleichbar mit den experimentellen Ergebnissen.

^{2,17}Die Erhaltung des normalen Gleichgewichtes der zellulären Funktionen gegenüber Störeinflüssen, was sich meist durch konstante intrazelluläre Konzentrationen an Enzymen, Metaboliten, etc. ausgedrückt.

Bei dem *Testen* eines Modells wird anhand von Testdatensätzen oder Testfällen überprüft, ob das Modell Fehler oder Ungenauigkeiten aufweist. Die Tests werden durchgeführt, um ein Modell zu validieren bzw. es zu verifizieren, so daß die Vorgänge der Verifikation, Validierung und des Testen meist nicht getrennt voneinander betrachtet werden.

Die VV&T Techniken wurden zum größten Teil in den Bereichen des „Software Engineering“ und der „Modellierung und Simulation“ entwickelt. Über 75 verschiedene Techniken sind in (Balci 1998) beschrieben und in vier Kategorien eingeteilt worden. Den informellen Techniken liegt kein strikter mathematischer Formalismus zugrunde, sondern sie verlassen sich auf das menschliche Urteilsvermögen, z.B. zur Prüfung durch Experten, ob die Ergebnisse eines Modells in gegebenem Zusammenhang sinnvoll erscheinen (Hermann 1967). Statische Techniken beschäftigen sich mit dem Modellentwurf und Quelltext, ohne daß das Modell dabei ausgeführt, d.h. simuliert werden muß. Zu ihnen gehören z.B. die symbolische Evaluation einer Funktion bzw. eines Sub-Modells, semantische und Syntaxanalysen, die u.a. durch den Compiler einer Simulationssprache durchgeführt werden, oder Analysen der Modellstruktur (Whitner und Balci 1989). Dynamische Techniken erfordern die Ausführung des Modells und untersuchen das Modellverhalten. Zu ihnen gehört das funktionale oder „Black-Box“-Testen (Howden 1980), bei dem Testdaten eingegeben werden und die Resultate des Modells für diese Eingangsgrößen evaluiert werden. Graphische Vergleiche zwischen den Simulationsergebnissen und dem realen Verhalten (Wright 1972), die Sensitivitätsanalyse (Miller 1974) und verschiedene statistische Tests stellen weitere der zahlreichen Möglichkeiten zur dynamischen VV&T dar. In Analogie zu dem Modellierungsprozeß gibt es dabei „Top-Down“-Strategien, bei denen zunächst ein Modell auf der höchsten Ebene entwickelt und durch Hinzunahme immer weiterer validierter Sub-Modelle verfeinert wird oder „Bottom-Up“-Strategien, bei denen der Prozeß umgekehrt erfolgt (Whitner und Balci 1989). Formale Techniken versuchen, die Korrektheit des Modells mathematisch durch Induktion, logische Deduktion oder Inferenz zu beweisen (Khanna 1991). Dies ist jedoch häufig für komplexe Modelle nicht vollständig möglich.

Die Entscheidung über die Gültigkeit eines Modells auf der Basis der vorgestellten Techniken erfolgt in der Regel von den Modellierern selbst im Verlauf des Modellierungsprozesses, oder von einer dritten, unabhängigen Seite während oder nach dem Prozeß (Wood 1986). In diesem Zusammenhang wird unter der Akkreditation eines Modells die Vergabe eines offiziellen Zertifikates verstanden, daß ein Modell oder eine Simulation annehmbar für den Gebrauch zu einem bestimmten Zweck ist.

2.3.2 Modellierungs- und Validierungsprozeß

In (Page 1991; Balci 1997) werden verschiedene Prinzipien diskutiert, auf denen ein konkreter Validierungsansatz für ein bestimmtes Simulationsmodell basieren sollte. Ein wesentliches Prinzip ist, daß die Verifikation und Validierung aus Gründen der Effizienz während des gesamten Modellierungsprozesses zu erfolgen hat, weil auf diese Weise die Qualitätsmängel des Modells in derjenigen Modellierungsphase identifiziert und berichtigt werden können, in der sie auftreten. Der Modellierungsprozeß kann auf unterschiedliche Art charakterisiert werden. Komplexe Darstellungsformen schließen z.B. Stufen für die Findung der geeigneten Modellierungstechnik, die Entwicklung einer geeigneten Form der Modellrepräsentation oder die experimentelle Versuchsplanung ein (Balci 1998). Eine einfachere Art der Darstellung, die im allgemeinen (Banks et al. 1995) und für diese Arbeit zur Verdeutlichung des Validierungskonzeptes ausreicht, ist in Abb. 2–4 gezeigt (Sargent 1982). Das reale System bezeichnet das zu modellierende Phänomen. Das konzeptuelle Modell wird durch Analyse des Systems und anschließende Modellierung als mathematische, logische oder verbale Repräsentation des realen Systems entwickelt. Das Computermodell entsteht über die Implementierung des konzeptuellen Modells und ist ausführbar, wobei die Simulations- mit experimentellen Ergebnissen verglichen werden können. Die

folgende Diskussion des den Modellierungsprozeß begleitenden Validierungsprozeß wird in Analogie zu (Page 1991; Sargent 2000; Sargent 2001) geführt.

Die *Validität der Daten* gewährleistet, daß die Daten, die für die Bildung und die Validierung des konzeptuellen Modells angemessen und richtig sind. Die Validität der Daten bildet somit die Grundvoraussetzung für die Modellierung und wird daher häufig nicht als Teilschritt der Modellvalidierung betrachtet. Wird die Validierung der Daten in den Validierungsprozeß integriert, so ist sie bereits während der Erstellung des konzeptuellen Modells durchzuführen, indem die Daten auf Plausibilität, Konsistenz bzw. Ausreißerwerte überprüft sowie in der Regel in geeigneter Form geglättet werden.

Die *Validierung des konzeptuellen Modells* untersucht, ob die Theorien und Annahmen, die dem konzeptuellen Modell zugrundeliegen, richtig sind, und ob die Repräsentation des realen Systems durch das Modell für eine gegebene Fragestellung sinnvoll ist. Dazu sollten die Hypothesen und vereinfachenden Annahmen, von denen der Modellentwurf ausgeht, mit bekannten Theorien des Anwendungsgebietes vereinbar sein und einer Plausibilitätsprüfung durch Fachleute mit einem Erfahrungswissen über das reale System standhalten. Wenn möglich, sollte durch den Vergleich mit empirischen Daten auf der Basis von statistischen bzw. mathematischen Methoden nachgewiesen werden, daß solche Annahmen keine allzu großen Abweichungen zum realen System zur Folge haben und damit vertretbar erscheinen. Weiterhin sollte der innere Aufbau des Modells auf seine Plausibilität untersucht werden. Dabei ist u.a. auch mit Fachleuten zu diskutieren, ob für das Modell – in Bezug auf die Fragestellung – die problemrelevanten Objekte, Attribute und Beziehungen korrekt erfaßt wurden^{2,18}. Eine notwendige Voraussetzung für die Diskussion mit den Experten ist die transparente Darstellung des Modells, z.B. die Auflistung aller dem Modell zugrundeliegenden Gleichungen oder eine graphische Repräsentation mit Strukturdiagrammen etc. Die *Verifikation des Computermodells* sichert ab, daß die Programmierung

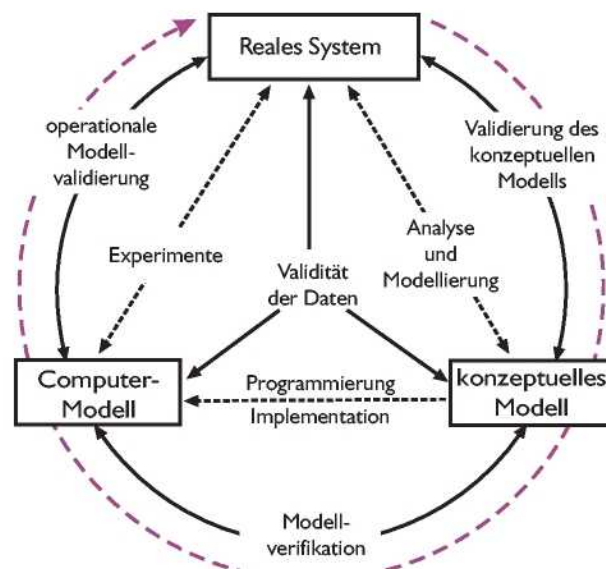


Abb. 2–4 Modellierungs- und Validierungsprozeß aus (Sargent 1982); die gestrichelte, äußere Linie gibt den Verlauf des Prozesses an

und Implementierung des konzeptuellen Modells korrekt erfolgt ist. Der Prozeß der Verifikation hängt wesentlich davon ab, ob eine spezielle Simulationssprache, wie R, Matlab, Maple, oder eine höhere Programmiersprache, wie Fortran, C, C++, Java, zur Implementierung benutzt wurde. Im ersten Fall ist

^{2,18} Werden in verschiedenen Phasen des Modellierungsprozeß Modellanwender und andere Kenner des realen Systems herangezogen, um das Modellverhalten zu interpretieren und zu beurteilen, spricht man von „face validation“.

das Fehlerrisiko, die für die Implementierung benötigte Zeit, aber auch die Flexibilität bei der Programmierung verringert. Die Verifikation beschränkt sich dann darauf zu untersuchen, ob die Sprache auf dem Computer richtig implementiert (bzw. installiert) und das Modell korrekt in der Sprache programmiert wurde. Wenn höhere Programmiersprachen verwendet werden, um eine spezielle Simulationssoftware zu erstellen, sollten bei der Entwicklung der Software Methoden des Software-Engineering, wie die schrittweise Programmentwicklung und -verfeinerung, Modularisierung und systematische Programmdokumentation verwendet werden. In diesem Fall ist zu verifizieren, ob die Simulationsfunktionen, wie z.B. Routinen zur Integration eines DGL-Systems, und das Computermodell korrekt implementiert wurden. Zu diesem Zweck dienen statische oder dynamische Programmtestmethoden, s. z.B. (Whitner und Balci 1989).

Während der aufwendigsten Phase des Validierungsprozesses, der *operationalen Modellvalidierung*, wird festgestellt, ob das Verhalten des Computermodells für den beabsichtigten Zweck und den Anwendungsbereich des Modells Gültigkeit besitzt. Hierbei ist es bei Verwendung von vergleichenden Validierungsansätzen wichtig, daß das reale System beobachtet werden kann, d.h. daß experimentelle Daten zur Untersuchung des Modellverhaltens erfaßt werden können (Sargent 2000). Ansonsten ist nur der Vergleich mit anderen Modellen möglich.

Eine Vielzahl von Graphen, in denen unterschiedliche Größen, wie der Mittelwert, die Varianz bzw. die Zeitreihe einer Variablen, Zusammenhänge zwischen zwei verschiedenen Meßgrößen einer Zustandsvariablen oder zwischen einer Meßgröße und einer Variablen des Modells aufgetragen sind, kann für eine subjektive Beurteilung im Sinne der „Face Validation“ oder in Turing Tests^{2,19} benutzt werden (Sargent 1996).

Eine andere Möglichkeit für die vergleichende Beurteilung des Modells bietet die Berechnung von Konfidenzintervallen für die Ergebnisdifferenz zwischen Simulationsdaten und empirischen Daten. Befindet sich der Wert Null nicht innerhalb des Konfidenzintervalls, so unterscheiden sich Modell und Realsystem hinsichtlich der betrachteten Ergebnisvariablen signifikant, d.h. das Modell sollte verworfen werden (Law und Kelton 1991). Für den Vergleich mehrerer Variablen werden simultane Konfidenzintervalle mit Hilfe von multivariaten, statistischen Methoden berechnet (Balci und Sargent 1984). Ähnlich zu der Berechnung von Konfidenzintervallen werden Hypothesentests benutzt, um Mittelwerte, Varianzen oder Zeitreihen von Modell- und Systemvariablen dahingehend zu vergleichen, ob das Systemverhalten durch das Modell mit hinreichender Genauigkeit wiedergegeben wird (Balci und Sargent 1982a; Balci und Sargent 1982b; Banks et al. 1995). Die Nullhypothese, H_0 : das Modell ist bei Annahme eines angemessenen Bereiches für die (Wiedergabe-) Genauigkeit und unter den gegebenen experimentellen Bedingungen gültig, wird in diesem Fall gegen die Hypothese, H_1 : das Modell ist nicht gültig, getestet. Dabei sind zwei Arten von Fehlern möglich. Infolge von Fehlern des Typs 1 (α) wird ein eigentlich gültiges Modell als inkorrekt abgelehnt (die Wahrscheinlichkeit, den Fehler zu begehen, kann als des Risiko des Modellierers bezeichnet werden (Sargent 2000)), bei Fehlern des Typs 2 (β) wird ein inkorrektes Modell als gültig akzeptiert (Risiko der Nutzer des Modells). Der Grad der Ähnlichkeit zwischen Modell und System wird durch eine – meist statistische – Prüfgröße λ ausgedrückt, die so gewählt wird, daß der Ähnlichkeitsgrad für ansteigende Werte von λ sinkt. Der akzeptable Bereich für die Genauigkeit der Wiedergabe des Systemverhaltens durch das Modell, $0 \leq \lambda \leq \lambda^*$ bestimmt den Bereich, innerhalb dessen das Modell als gültig erachtet werden kann.

Im Gegensatz zu den vergleichenden Verfahren der operationalen Modellvalidierung berücksichtigen explorative Validierungsansätze das Systemverhalten nicht sondern untersuchen nur das Modellverhalten. Die am häufigsten verwendete Methode ist hierbei die Sensitivitätsanalyse (Page 1991), bei der

^{2,19}Fachleute, die mit dem modellierten Realsystem gut vertraut sind, werden vor die Aufgabe gestellt, die Modellergebnisse und die realen Daten zu unterscheiden. Sind sie dazu nicht in der Lage, hat das Modell diesen Test bestanden.

bestimmt wird, wie die Modellausgabe auf Veränderungen der Modelleingabe oder der Modellstruktur reagiert, und von welchen exogenen Variablen, Relationen oder Parametern die Modellergebnisse am stärksten abhängen. Dadurch ist es möglich, eventuelle Strukturfehler im Modell aufzudecken, wesentliche Einflußgrößen von unwesentlichen, evtl. zu vernachlässigenden, Variablen zu trennen und Aussagen über die Stabilität des Modells zu gewinnen. Außerdem läßt sich herausfinden, ob bestimmte Modellresultate nur für eingeschränkte Wertebereiche von Parametern gültig sind.

Im Zuge des gesamten Modellierungs- und Validierungsprozesses werden in der Regel verschiedene Modelle oder mehrere Versionen eines Modells entwickelt, bevor ein für den jeweiligen Anwendungszweck ausreichendes Modell erhalten werden kann.

3 Konzept für den Modellierungs- und Validierungsprozeß

3.1 Einleitung

Wie in Abschnitt 2.3 dargestellt, sollte der Modellierungs- und Validierungsprozeß (MVP) auf die jeweils aktuelle Fragestellung abgestimmt werden, um möglichst sinnvolle Lösungen zu erhalten. Das erarbeitete Konzept für einen Prozeß, der zur Beantwortung der Fragestellung: „die Auswertung von Datensätzen aus Substrat-Puls-Experimenten mit Modellen, aus denen Aussagen für die genetische Modifikation von Mikroorganismen im Sinne des Metabolic Engineering abgeleitet werden können“, geeignet ist, wird in den folgenden Abschnitten umrissen.

Das reale System ist der Stoffwechsel des Mikroorganismus *Escherichia coli*. Spezielle Stoffwechselabschnitte, s. Anhang A, sind unabhängig von dieser Arbeit mit Substrat-Puls-Experimenten, s. Abschnitt 3.2, untersucht worden, die fünf aufeinanderfolgende, zeitintensive Stufen durchlaufen. Die Gesamtdauer des Experimentes beträgt 3-4 Monate^{3.1}, wobei den eigentlichen Experimenten eine Zeitspanne von 1-2 Jahren für die Institution des jeweiligen Versuchsaufbaus bzw. für die Einarbeitungsphase (Schaefer 1999; Buchholz 2001; Oldiges 2004) vorausgeht. Dies bedeutet, daß die Reproduktion der Experimente im Zeitrahmen einer Dissertation mit praktischer bzw. die Rückkopplung von den Modellen auf die Experimente im Rahmen einer Dissertation mit theoretischer Zielstellung nicht ohne weiteres möglich ist. Damit stellt sich für diese Arbeit in erster Linie das Ziel, den vereinfachten, in Abb. 2–4 gezeigten MVP einmalig zu durchlaufen, wobei besonderer Wert auf die Validierung der Daten, s. Abschnitt 3.3.1, und die operationale Modellvalidierung, s. Abschnitt 3.4.2 gelegt werden sollte.

Da während der Experimente zeitabhängige Meßgrößen aufgenommen wurden und ein hohes Maß an quantitativer Information in Form von enzymkinetischen Gleichungen bzw. Parametern für die untersuchten Stoffwechselabschnitte vorliegt, kann das konzeptuelle Modell unter Berücksichtigung der in Abschnitt 3.3 geschilderten vereinfachenden Annahmen durch ein System gewöhnlicher Differentialgleichungen erster Art wiedergegeben werden. Zur Implementierung des DGL-Systems waren in einer Vorläufer- bzw. einer parallel durchgeführten Arbeit die Softwarewerkzeuge MMT, Metabolic Modeling Tool (Hurlebaus 2001; Hurlebaus et al. 2002), bzw. MMT2 (Haunschild 2001) erstellt worden. Sie wurden im Rahmen dieser Arbeit zum ersten Mal verwendet, getestet und im Fall von MMT weiterentwickelt, s. Abschnitt 3.4.

3.2 Substrat-Puls-Experimente

Messungen der intrazellulären Metabolitkonzentrationen wurden erstmalig im Jahr 1973 durchgeführt, um Erkenntnisse über Stoffwechselvorgänge in der lebenden Zelle zu erhalten (Gancedo und Gancedo 1973). Die Quantifizierung intrazellulärer Konzentrationen im Anschluß an die Aufgabe von Substrat-Pulsen auf unter Substrat-Limitierung kultivierte Mikroorganismen stellt seit Anfang der 90'er Jahre eine häufig verwendete Methode zur experimentellen Verifikation von Stoffwechselmodellen dar (de Koning und van Dam 1992; Theobald et al. 1993; Theobald 1995; Rizzi et al. 1996; Theobald et al. 1997b; Weuster-Botz 1997; Mailinger et al. 1998; Schaefer et al. 1999; Vaseghi et al. 1999; Lange et al. 2001; Ostergaard et al. 2001; Vaseghi et al. 2001; Visser 2001; Buziol et al. 2002; Visser et al. 2002). Als Beobachtungszeitraum für den transienten Zustand eignet sich aufgrund der hohen Umsatzraten der Stoffwechselprodukte nur ein Zeitfenster von Sekunden, um die schnellen Vorgänge der metabolischen

^{3.1}Wiederholte mündliche Mitteilungen von A. Buchholz, S. Zaja, M. Oldiges, IBT-2, Forschungszentrum Jülich, Zeitraum von Oktober 2000 bis Mai 2003.

Regulation zu erfassen und eine von genetischen Regulationsvorgängen unbeeinträchtigte, nahezu konstante Konzentration der Enzyme zu gewährleisten.

Zur Durchführung der invasiven Experimente werden Methoden für eine schnelle, reproduzierbare Probennahme im Bereich von Millisekunden, für einen möglichst vollständigen Zellaufschluß unter rascher Inaktivierung der Enzyme und für eine schonende Extraktion der Metabolite sowie für die Bestimmung von Metabolitkonzentrationen im mikro-molaren Bereich benötigt.

Die dieser Arbeit zugrundeliegenden Daten stammen zum einen aus einem Substrat-Puls-Experiment mit einem *Escherichia coli* Wildstamm (Experiment WT1) (Buchholz 2001). Dieses wurde schon zu Beginn der Arbeit durchgeführt und bildet somit die Grundlage für den Test der entwickelten Methoden zur Modellbildung und Validierung. Zum anderen existieren zwei verschiedene Meßdatensätze aus Experimenten mit L-Phenylalanin-Produktionsstämmen (Experimente MU1, MU2) (Oldiges 2004), die erst gegen Ende der Arbeit vorlagen und punktuell für detailliertere Analysen genutzt werden konnten. Die Versuchsdurchführung der unterschiedlichen Experimente wird im Folgenden dargestellt.

1. Stämme und Kultivierung

Die Experimente WT1, MU1 und MU2 wurden jeweils einmal durchgeführt, wobei die Organismen in allen Fällen unter Limitierung von Glucose kultiviert wurden, s. Abschnitte 3.2.1 und 3.2.2 für die Charakteristika der Stämme und ihrer Fermentationen.

2. Pulsaufgabe

Ein Puls-Aufgabe-System ermöglichte die Aufgabe des Glucose-Pulses in den Bioreaktor (Buchholz 2001; Oldiges 2004). Die Auslösung des Pulses, d.h. die Öffnung der zwei Pulsventile, erfolgte während des automatischen Wechsel des ersten und des zweiten Magazins (s. Punkt 3). Die Glucose-Lösung (70 ml einer Konzentration von $2,33 \text{ mol l}^{-1}$) wurde nahezu ohne Druckabfall in den Fermenter injiziert, so daß mit einer Mischzeit, $\Theta_{90} = 0,6 \pm 0,2 \text{ s}$ in WT1 bzw. $\Theta_{90} \approx 0,8 \text{ s}$ in MU1 und MU2^{3.2} ein Glucose-Überschuß von ca. 3 g l^{-1} im Fermenter resultierte.

3. Probennahme

Die automatisierte Probennahme (Paschold et al. 1998) gestattet eine kontinuierliche Probennahme in einem zeitlichen Abstand von 220 ms, wobei die Zellsuspension aus einem am Boden des Reaktors befindlichen Probennahmeventil in die Probennahmegefäße gelangt, die in Magazinen mit je 16 Probenhaltern auf einem Förderband unter dem Ventil vorbei bewegt werden (Schaefer et al. 1999). Insgesamt werden 160 Proben entnommen (16 vor dem Puls, 144 nach dem Puls).

Während der Probennahme wurden zur raschen Inaktivierung des Stoffwechsels 5 ml (WT1, MU2), bzw. 7,5 ml (MU1) der 37°C warmen Zellsuspension in 15 ml einer ca. -50°C kalten wäßrigen 60% Methanol-HEPES-Lösung eingespritzt. Die Gefäße wurden nach der Probennahme manuell verschlossen und lagerten bei -28°C.

4. Zellaufschluß

Nach 5-minütiger Zentrifugation der Proben bei -20°C und 10.300 g wurde der Zellrückstand mit wäßriger 60% Methanol-HEPES (70 mM) Lösung re-suspendiert und anschließend mit -18°C kalter 43,8% (w/v) Perchlorsäure aufgeschlossen. Im Anschluß wurden die Proben mit einem Vortex Schüttelgerät homogenisiert und bei -80°C für 1-2 Tage eingefroren. Die aufgetauten Proben wurden bei 4°C für 30 Minuten ultrazentrifugiert (42.000 g), um die Zelltrümmer und Proteine abzutrennen. Der klare Überstand wurde mit gesättigter K_2CO_3 -Lösung manuell neutralisiert. Das entstandene KClO_4 wurde durch eine erneute Zentrifugation bei 4°C und 42.000 g für 5 Minuten

^{3.2} Θ_{90} entspricht der Zeit, die notwendig ist, um die stoßförmige Zugabe (entspricht im Idealfall dem Dirac-Puls und in der Praxis der Delta-Funktion) eines Spurstoffs im Reaktor bis zu einem definierten Mischungsgrad von 90% räumlich zu homogenisieren.

abgetrennt. Die neutralen Zellextrakte lagerten bis zur quantitativen Bestimmung bei -28°C .

5. Analytische Methoden zur Quantifizierung der Metabolite

Die Quantifizierung der Metabolite aus der Glykolyse erfolgte in WT1 über gekoppelte enzymatische Assays (Buchholz 1998; Schaefer 1999), während die Nukleotide und die Metabolite aus dem Pentose-Phosphat-Weg über die ESI-LC-MS-Methode (ESI: Elektrospray-Ionisation) (Buchholz et al. 2001) quantifiziert wurden, s. Anhang C für einen Überblick über die mit den verschiedenen Methoden erfaßten Metabolite.

Alle in MU1 und MU2 erfaßten Metabolite wurden mit einem Triple Quadrupol LC-MS Gerät der Firma Thermofinnigan quantifiziert (Oldiges 2004), wobei die in (Buchholz 2001) beschriebene HPLC-Methode im Wesentlichen beibehalten wurde (HPLC: High-Pressure-Liquid-Chromatography). Die intrazellulären Metabolitkonzentrationen, $\tilde{c}$, in einer Probe werden dann aus der über die analytische Methode quantifizierten Konzentration, $\tilde{c}_{\text{LCMS}}$, berechnet, s. (Buchholz 2001):

$$\begin{aligned}\tilde{c} &= \tilde{c}_{\text{LCMS}} \cdot F_{\text{verd, LCMS}} \cdot F_{\text{verd, Extr.}} \\ F_{\text{verd, LCMS}} &= \frac{V_{\text{Probe}} + V_{\text{Puffer}}}{V_{\text{Probe}}} \\ F_{\text{verd, Extr.}} &= \frac{BTM \cdot V_S \cdot K_E + V_E + V_N + V_X}{BTM \cdot V_S \cdot K_E},\end{aligned}\tag{3-1}$$

$F_{\text{verd, LCMS}}$: Verdünnungsfaktor der LC-MS-Bestimmung; $F_{\text{verd, Extr.}}$: Verdünnungsfaktor des Extraktionsvorgangs; V_{Probe} : Volumen der Probenlösung für die LC-MS-Bestimmung; V_{Puffer} : Volumen der Pufferlösung für die LC-MS-Bestimmung; BTM : Konzentration der Biotrockenmasse im Fermenter; V_S : dem Fermenter entnommenes Probenvolumen; K_E : spezifisches Zellvolumen; V_E : Volumen an Perchlorsäure und Resuspensionslösung für die Zellextraktion; V_N : Neutralisationsvolumen; V_X : Flüssigkeitsvolumen, das nach Zentrifugation der Zellsuspension und anschließender Entfernung des Fermentationsüberstand noch im Zellpellet verbleibt.

3.2.1 Experiment mit einem *Escherichia coli* Wildstamm

Die Versuche erfolgten mit dem Stamm *E.coli* K12 DSM 498 (DSM Stammhaltung D), einem Wildtypstamm des Bakteriums (Buchholz 2001).

Die Organismen wurden kontinuierlich unter Glucose-Limitierung kultiviert, s. Anhang B für die verfahrenstechnischen Grundlagen. Dazu wurden die Vorkulturen nach Erreichen von $\text{OD}_{650} \approx (3-4)$ in die Hauptkultur inokuliert, s. (Schaefer 1999; Buchholz 2001) für die Medienzusammensetzung. Nach einer anfänglichen Batch-Phase wurde bei Erreichen einer optischen Dichte von $\text{OD}_{650} \approx 15$ die Medienregulierung und die Gewichtsdosierung aktiviert, um eine Verdünnungsrate von $D=0,125 \text{ h}^{-1}$ einzuhalten. Über die Regelung der Nährstoffzufuhr, des pH-Wertes auf 7 und der Temperatur auf 37°C wurde ein Fließgleichgewichtszustand erreicht, der sich in konstanten Werten für die Glucose-Konzentration (unterhalb des Detektionslimits), die Biotrockenmasse ($BTM = 10 \text{ g l}^{-1}$), den Sauerstoffpartialdruck ($p_{\text{O}_2}=(82 \pm 2)\%$) und den CO_2 -Partialdruck ($p_{\text{CO}_2}=(30 \pm 2)\%$) ausdrückte. Der Fließgleichgewichtszustand wurde drei bis vier Verweilzeiten vor Injektion des Glucose-Pulses aufrechterhalten.

3.2.2 Experimente mit Phenylalanin-Produktionsstämmen

Basierend auf *E.coli* K12 W3110 wurde der genetisch veränderte Stamm *E.coli* 4 als L-Phenylalanin-(Phe-) Produktionsstamm entwickelt. Um eine die Raum-Zeit-Ausbeute beeinträchtigende Bildung des

Nebenproduktes L-Tyrosin und eine übermäßige genetische Regulation durch L-Phenylalanin zu vermeiden, wurden sowohl das Gen *tyrA* (Chorismat-Mutase / Prephenat-Dehydrogenase) als auch die *pheA* (Chorismat-Mutase / Prephenat-Dehydratase) und *aroF* (Tyrosin-spezifische 3-Deoxy-D-arabino-heptulosonat-7-phosphat-Synthase oder DAHP-Synthase (Tyr)) chromosomal deletiert. *E.coli* 4 ist somit L-Tyrosin- und zunächst auch L-Phenylalanin-auxotroph.

Zur L-Phenylalanin-Produktion wurde in *E.coli* 4 das Plasmid pF20 transformiert. Es basiert auf dem Expressionsvektor pJF119EH und besitzt eine Ampicillinresistenz als Selektionsmarker. Weiterhin ist in dem Vektor ein durch Isopropylthio- β -D-galactosid (IPTG) induzierbarer, synthetischer *tac*-Promotor enthalten, der Glucose-resistent ist. Die in den Vektor klonierten Gene *aroF^{fbr}* und *pheA^{fbr}* kodieren die nicht von Tyr bzw. Phe inhibierten Enzyme 3-Deoxy-D-arabino-heptulosonat-7-phosphat-Synthase (Tyr) bzw. Chorismat-Mutase / Prephenat-Dehydratase. Damit ist die vorherige Regulation der spezifischen Enzymaktivitäten durch die Endprodukte des Biosyntheseweges, die sogenannte „Feedback-Inhibierung“, aufgehoben (*fbr*: Feedback-Resistenz). Aufgrund der Überexpression der Gene auf dem Klonierungsvektor sind gleichzeitig etwaige Regulationsmechanismen auf der Ebene der Transkription ausgeschaltet. Dieser sogenannte *E.coli* 4pF20-Stamm wurde für das Experiment MU1 verwendet.

Der in Experiment MU2 eingesetzte Stamm basiert auf *E.coli* 4pF20. Zusätzlich wurde das Gen *aroB^{wt}*, das für die 3-Dehydrochinat-Synthase kodiert, dem Plasmid 4pF20 hinzugefügt, da die niedrige Umsetzungsgeschwindigkeit dieses Enzyms die Phe-Produktivität in vorherigen Experimenten limitierte, s. (Gerigk 2001). In Tabelle 3–1 sind die verwendeten Stämme sowie die zugehörigen Experimente zusammengetragen.

Tab. 3–1 Übersicht der verwendeten Mikroorganismen

Name	Name, Genotyp	Versuch
LJ110	<i>E.coli</i> K12 W3110	WT1
4pF20	LJ110, $\Delta(pheA\ tyrA\ aroF)$, pJF119EH <i>aroF^{fbr}</i> <i>pheA^{fbr}</i> (amp ^R)	MU1
4pF78	LJ110, $\Delta(pheA\ tyrA\ aroF)$, pJF119EH <i>aroF^{fbr}</i> <i>aroB^{wt}</i> <i>pheA^{fbr}</i> (amp ^R)	MU2

Bei einer über eine Woche andauernden kontinuierlichen Fermentation wäre der Verlust der Plasmide infolge des großen Wachstums- und Selektionsdruckes in der Kultur zu hoch, so daß hier als Kultivierungsmethode nur eine zwei- bis dreitägige Fed-Batch-Fermentation in Frage kam. Die Vorkulturen wurden in der exponentiellen Wachstumsphase (ca. OD₆₅₀=(3-4) nach 13 h) auf das Tyr- und Ampicillin-haltige Fermentermedium (Gerigk 2001; Hirsch 2001) überimpft. Dadurch wurde eine lag-Phase (ca. 3 h) im Hauptkulturmedium erzielt. In der exponentiellen Wachstumsphase wurde eine mittlere spezifische Wachstumsrate von ca. 0,4 l / h erreicht. Die Sauerstoffkonzentration wurde über die Regelung der Rührer- und der Zuluftgeschwindigkeit dem Sauerstoffbedarf der Zellen während des Wachstums angepaßt, so daß ein charakteristisches Sägezahnprofil resultierte. Der pH wurde durch die Zugabe einer 25%-igen NH₃-Lösung während der gesamten Fermentation auf einen Wert von 6,5 eingeregelt.

Nach ca. 5-6 h waren die anfänglich vorgelegten Mengen an Glucose und L-Tyrosin im Fermenter fast vollständig verbraucht. Aus diesem Grund wurden gleichzeitig mit der IPTG-Induktion nach ca. 5,9 h (entspricht OD₆₅₀ $\approx$ 10-12) die Zulaufmedien Glucose und Tyr in den Fermenter eingeleitet, um das Zellwachstum weiterhin zu ermöglichen. Da eine Tyrosin-Limitierung während des Glucose-Pulses vermieden werden sollte, wurde der Zulauf an Tyr in den Fermenter erst kurz vor dem Puls abgestellt. Damit lagen in beiden Experimenten noch signifikante Tyr-Konzentrationen von ca. 1,4 mM im Bioreaktor vor, und das Erreichen der stationären Phase sollte damit allein über die Steuerung des Glucose-Zulaufs erreicht werden. Hierbei wurde die Annahme zugrundegelegt, daß sich die intrazellulären Metabolite

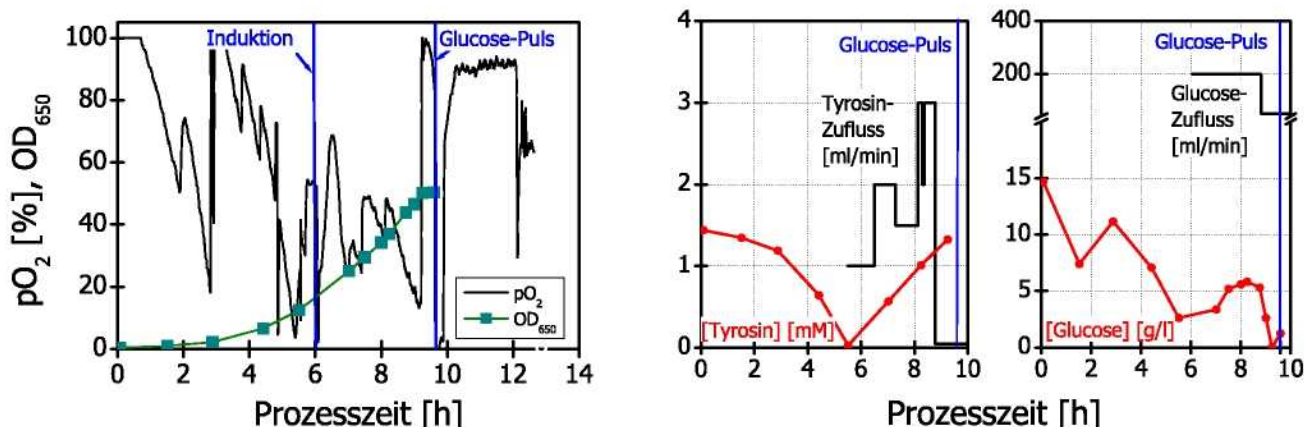


Abb. 3-1 Fermentationsverlauf für die Hauptkultur von MU1; links: Verläufe der Zelldichte und des Sauerstoffpartialdrucks im Reaktor, rechts: Steuerung des Zulaufs von Glucose und Tyrosin in den Reaktor bzw. Konzentrationen der Spezies im Reaktor

aufgrund der hohen Stoffwechselaktivität der Zellen nach ca. 20 bis 30 Minuten in einem Fließgleichgewichtszustand befinden^{3,3}. In beiden Experimenten wurde daher bei einer Biomassekonzentration von $OD_{650} \approx 45$ (entspricht ca. 8,8 h bei MU1 und ca. 10,1 h bei MU2), der Glucose-Zulauf in den Fermenter abgestellt und nach ca. 30 Minuten der Glucose-Puls injiziert, s. Abb. 3-1.

3.2.3 Datensätze

Abbildung 3-2 zeigt beispielhaft die Trajektorien zweier Metabolite der Glykolyse für den Vergleich der drei Substrat-Puls-Experimente. Die Trajektorien aller in den Experimenten erfaßten intrazellulären Metabolite sind in Anhang C zusammengestellt. Mit Ausnahme der Glucose-Konzentrationen in den Experimenten WT1 und MU2 wurden die extrazellulären Metabolitkonzentrationen nicht quantifiziert.

Allen Daten ist gemein, daß die intrazellulären Metabolitkonzentrationen hochfrequenten Schwankungen unterliegen. Vor Beginn der Modellierung ist es daher wesentlich zu klären, wie die Streuungen in dem Verlauf aller Trajektorien zu deuten sind. Sind sie durch Reaktionen des Stoffwechsels verursacht, so müßte die Frequenz der Probennahme noch erhöht werden, um aussagefähige Modelle zu bilden. Wird, wie in dieser Arbeit, angenommen, daß das Rauschen während der Versuchsdurchführung, z.B. durch den Zellaufschluß oder durch die Neutralisation der aufgeschlossenen Lösungen, verursacht wird, so sollten die Daten für die Modellierung in geeigneter Weise aufbereitet werden, s. Abschnitt 3.3 bzw. 4.5 zur Untersuchung verschiedener Methoden zur Datenaufbereitung.

^{3,3}Mündliche Mitteilung, M. Oldiges, 16.5.2003

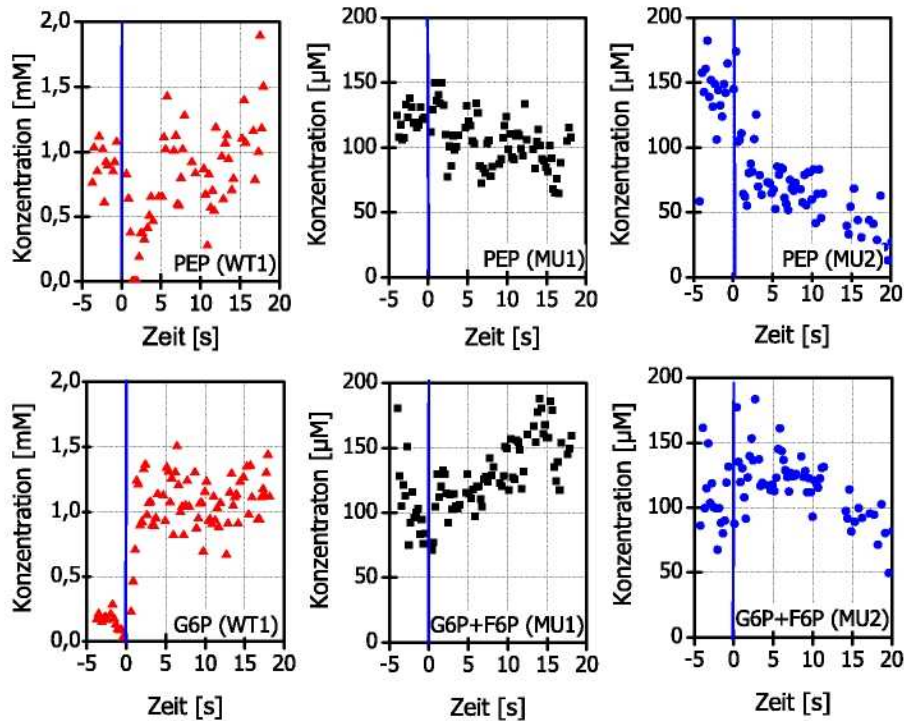


Abb. 3–2 Die in den Experimenten WT1 (links), MU1 (mitte) und MU2 (rechts) bestimmten Verläufe der Konzentrationen von Phosphoenolpyruvat (PEP) und Glucose-6-Phosphat (G6P); die senkrechten Linien bei 0 s zeigen den Zeitpunkt des Pulses an; in MU1 und MU2 konnte nur die Summe der Konzentrationen von G6P und F6P (Fructose-6-Phosphat) erfaßt werden, da die Metabolite isobar sind und nicht über die HPLC-Säule aufgetrennt werden konnten

3.3 Konzeptuelles Modell

3.3.1 Validierung der Daten

Da die Validität der Daten eine notwendige Voraussetzung für den gesamten MVP bildet, s. Abschnitt 2.3.2, sind die experimentellen Datensätze vor Beginn der Modellierung zu überprüfen. Die folgenden Charakteristika der Daten und Experimente waren ausschlaggebend für die Wahl der Methoden zur Datenvalidierung:

1. Die Experimente werden in der Regel nicht wiederholt. Dies hat zur Folge, daß etwaige Meßfehler, d.h. Ausreißerwerte oder innerhalb eines Zeitabschnitts inkonsistent verlaufende Trajektorien einiger Metabolite, experimentell nicht nachgewiesen werden können.
2. Die Durchführung der Experimente ist komplex, d.h. zeit- und arbeitsintensiv. Das kann leicht dazu führen, daß evtl. nicht alle wesentlichen Größen in der jeweils erforderlichen Genauigkeit erfaßt und eindeutig protokolliert werden.
3. Die komplizierte Berechnung der intrazellulären Konzentrationen aus den Meßsignalen der Analysenmethode erfolgt nicht-automatisiert. Dadurch erfordert dieser Arbeitsschritt einen relativ hohen Zeitaufwand und eine gewisse Wahrscheinlichkeit für Rechenfehler ist gegeben.
4. Die Verläufe der intrazellulären Metabolitkonzentrationen weisen hochfrequente Schwankungen auf.
5. Eine uneinheitliche Aufnahme und Auswertung der Meßsignale beeinträchtigt die Transparenz der Experimente. Daher sollte die Protokollierung standardisiert erfolgen, um eine einheitliche,

übersichtliche und vollständige Darstellung der Meßdatenerfassung zu gewährleisten. Bei der Protokollierung sollten auch für alle Experimente gleiche Kriterien (Standards) angegeben werden, aufgrund derer eine Beurteilung der Datenqualität – z.B. für den Vergleich verschiedener Trajektorien eines Experimentes oder verschiedener Experimente – möglich ist.

Eine Fehlerrechnung stellt im Allgemeinen das Mittel der Wahl für die Erkennung etwaiger Ausreißer (Punkt 1) und die Beurteilung der Datenqualität (Punkt 5), z.B. auch der Schwankungen der intrazellulären Konzentrationen (Punkt 4), dar. Hier wird sie darüber hinaus eingesetzt, um die Genauigkeit, mit der die einzelnen Meßgrößen angegeben werden sollten (Punkt 2), zu ermitteln, s. Abschnitt 4.2.

Die Bedeutung der Standardisierung der Versuchsdurchführung und ihrer Protokollierung sowie die Bedeutung einheitlicher Kriterien zur Bewertung von Meßdaten (Punkt 5) wird besonders im Kontext der Systembiologie häufig erwähnt (Kitano 2002b; Alliance for Cellular Signaling (USA) 2004; Systeme des Lebens, Systembiologie (D) 2004). Im Gegensatz zu den meisten analytischen Methoden im Bereich der Transkriptomics und Proteomics, für die jeweils spezielle Standards entwickelt und noch kontrovers diskutiert werden (Roux-Rouquié et al. 2003), liegen einheitliche Kriterien zur Bewertung der hier verwendeten Methoden, mit denen niedermolekulare chemische Verbindungen seit Jahren quantifiziert werden^{3,4}, schon vor und sind als Bestimmungs-, Nachweis- und Erfassungsgrenzen im Rahmen von DIN-Vorschriften definiert (DIN: Deutsches Institut für Normierung). Diese werden neben anderen statistischen Größen auch in dieser Arbeit zur Beurteilung der Datenqualität herangezogen, s. Abschnitt 4.3.

Zur Standardisierung der Protokollierung (Punkt 5) und zur vereinfachten, automatisierten Berechnung der Meßdaten, der Bestimmungs-, Nachweis- und Erfassungsgrenzen und der Meßunsicherheiten (Punkt 2) wurde im Rahmen dieser Arbeit ein spezielles VBA^{3,5}-Programm entwickelt, s. Abschnitt 4.4. Die experimentell erfaßten Größen können in vorgefertigte Tabellenblätter eingetragen, abgespeichert und ausgewertet werden, was die übersichtliche, einheitliche und vollständige Protokollierung gewährleistet. Aufgrund der zeitsparenden Durchführung der Rechenoperationen ist eine Auswertung der Meßsignale bereits während ihrer Erfassung möglich, so daß Hinweise auf evtl. nochmals durchzuführende Messungen erhalten werden können.

Die innere Konsistenz der Daten (Punkt 1) kann über Plausibilitätstests geprüft werden, s. Abschnitt 2.3.2. Es bietet sich z.B. an, die Größenordnungen der Werte für die intrazellulären Metabolitkonzentrationen mit Literaturwerten zu vergleichen, oder zu testen, ob der Verlauf einzelner Trajektorien auf der Basis von gängigen Wissen über die Stoffwechselvorgänge erklärbar ist. Aufgrund der Komplexität des Stoffwechsels werden solche Tests jedoch nicht in jedem Fall eindeutige Ergebnisse liefern, so daß eine weitere Methode den Konsistenztest ergänzen sollte.

Hierfür wurde eine besondere Art der Zeitreihenanalyse^{3,6} gewählt, bei der Reaktionspfade aus der Korrelationsmatrix der Meßdaten rekonstruiert werden (Arkin und Ross 1995; Arkin et al. 1997). Sie wird zur Auswertung von Datensätzen aus dem Metabolom fast ausschließlich^{3,7} eingesetzt (Ross und

^{3,4}Eine quantitative Methode ist ein chemisches Verfahren, bei dem es um die Beantwortung der Frage geht, wie viel von einem Stoff in einer gegebenen Probe vorhanden ist. Zur jeweiligen Methode gehört u.a. die Wahl des Meßgerätes und des Meßverfahrens, z.B. Standardadditionsverfahren etc..

^{3,5}VBA: Visual Basic for Application.

^{3,6}Die Zeitreihenanalyse untersucht die Eigenschaften von Zeitreihen mit Hilfe von mathematisch-statistischen Methoden, insbesondere um Zusammenhänge zwischen aufeinanderfolgenden Beobachtungen zu erkennen und zukünftige Beobachtungen zu prognostizieren.

^{3,7}Die einzige hier bekannte Alternativmethode stellt die Ableitung des Reaktionsnetzwerkes aus der Jacobi-Matrix der Meßdaten dar (Chevalier et al. 1993; De la Fuente et al. 2001), deren erfolgreiche Anwendung auf *in-silico*, d.h. künstlich im Rechner erzeugte, Datensätze, erst Ende 2003 veröffentlicht (Hoops et al. 2003) und daher hier nicht eingesetzt wurde. Bei weiteren Methoden, wie (Vance et al. 2002; Torralba et al. 2003), bzw. Methoden des „Reverse Engineering“ genetischer Netzwerke (Wagner 2001), s. Abschnitt 2.2, sind die experimentellen Bedingungen oder die Beobachtungen nicht mit denen der Puls-Experimente vergleichbar.

Vlad 1999; Samoilov et al. 2001; Kauffman et al. 2002; Ross 2003), weil der Umfang ihrer Vorhersagen den der meisten anderen „Reverse Engineering“-Methoden übertrifft. Aufgrund der Ähnlichkeit der zeitlichen Konzentrationsverläufe können nämlich Aussagen darüber abgeleitet werden, ob die verschiedenen Spezies über direkte Reaktionen untereinander eng miteinander verwandt sind und wie die zeitliche Abfolge der Reaktionen ist, anstatt nur verschiedene Gruppen bzw. Cluster aus einander ähnlichen Metaboliten zu identifizieren. Das mit dieser Methode induktiv abgeleitete Reaktionsnetzwerk – d.h. die Meßdaten werden ohne Hinzunahme von *a-priori*-Wissen analysiert – kann mit den bekannten, unabhängig erstellten biochemischen Stoffwechseldiagrammen, s. Anhang A, verglichen und somit auf Inkonsistenzen geprüft werden. Zusätzlich wurde in dieser Arbeit das den Meßdaten überlagerte Rauschen mit diesem Verfahren genauer untersucht (Punkt 4). In Abschnitt 4.7 werden Potential und Grenzen der Methode für die Datenvalidierung evaluiert und ihre Anwendung auf verschiedene Datensätze demonstriert.

Um Punkt 4 zu klären, können in der Literatur erwähnte, unter ähnlichen Bedingungen durchgeführte Experimente zum Vergleich herangezogen werden. Obwohl die Frequenz der Probennahme solcher Experimente in der Regel niedriger als (4-5) Hz ist (s. Abschnitt 3.2), wurde dennoch z.B. aus den reproduzierten, relativ glatt verlaufenden Trajektorien in (Buziol et al. 2002) geschlossen, daß die Streuungen in den vorliegenden Meßdaten vermutlich nur zu einem sehr geringen Teil aus in der Bakterienzelle tatsächlich stattfindenden Vorgängen resultieren. Weiterhin wurden in einigen Fällen Schwierigkeiten bei der Berechnung der Modelle beobachtet, wenn die Rohdaten ohne nachfolgende Datenaufbereitung zugrundegelegt wurden, s. Abschnitt 3.4. Daher wurden verschiedene Methoden zur Filterung und Glättung der Datensätze geprüft, s. Abschnitt 4.5 und (Froemel 2003).

Alle erwähnten Methoden wurden unabhängig voneinander angewendet, um durch den Vergleich der verschiedenen Ergebnisse möglichst sichere Aussagen über die Meßdaten zu gewinnen.

3.3.2 Hypothesen und vereinfachende Annahmen

Auf der Basis von generellen, vereinfachenden Annahmen erschienen gewöhnliche DGL-Systeme erster Art als geeignet für die Beschreibung der experimentell bestimmten Zeitreihen. Sie werden im Folgenden dargestellt:

1. Eine Population von Mikroorganismen besteht aus Zellen, die sich hauptsächlich hinsichtlich Größe, Gewicht, Alter, chemischer Zusammensetzung und Zustand der genetischen Regulation unterscheiden und somit auch unterschiedliche Stoffwechselaktivitäten besitzen. Es gibt mehrere Modelle, die die Variationen der Population berücksichtigen, s. z.B. (Zangirolami et al. 1997; Chain und Chau 1998; Mantzaris et al. 1999). Für diese Arbeit werden sie jedoch vernachlässigt, da in kleineren Bioreaktoren kultivierte Mikroorganismen nur innerhalb geringer Größenbereiche variieren (Buchholz 2001), mit Substrat-Puls-Experimenten nur die chemische Komposition einer Durchschnittszelle untersucht werden kann und innerhalb eines dem Puls folgenden Zeitraums von 20-40 s keine nennenswerte Veränderung der Zelleigenschaften erwartet wird.
2. In einem Fermenter existieren Druck-, pH-, Temperatur- oder Nährstoffgradienten, wobei sich die Stoffwechselaktivität der Mikroorganismen in Abhängigkeit von den inhomogenen äußeren Bedingungen ändern kann. Diesbezüglich exakte Modelle berücksichtigen die Bewegung einzelner Zellen innerhalb des Fermenters, indem die Bewegung als stochastischer Prozeß modelliert wird (Kreft et al. 1998), oder der gesamte Raum in kleinere 3D-Räume zerlegt wird, die sich gegeneinander bewegen können (Hewitt et al. 2000). Für unseren Fall wird angenommen, daß aufgrund des kleinen Volumens des verwendeten Bioreaktors und der relativ kurzen Mischzeiten, s. Abschnitt 3.2, die räumlichen Gradienten vernachlässigt werden können.

3. Der Unterschied zwischen strukturierten und unstrukturierten Modellen zur Beschreibung von Zellen wurde bereits in Abschnitt 2.2.3 besprochen. Das erwähnte Modell zur Beschreibung von Substrat-Puls-Experimenten mit Hefezellen (Theobald et al. 1997a) berücksichtigt z.B. die unterschiedlichen Konzentrationen von NADH und ATP im Cytosol und in den Mitochondrien über jeweils zwei abhängige Variablen, die die zeitliche Veränderung der Komponenten in den beiden Kompartimenten beschreiben. Der Innenraum einer prokaryotischen Zelle, wie *E.coli*, ist nicht derartig kompartmentiert, so daß die hier erstellten Modelle nur insofern strukturiert sind, als sie von einem intra- und einem extrazellulären Raum ausgehen^{3.8}.
4. Es ist bekannt, daß ebenso wie im Bioreaktor auch Gradienten innerhalb der Zelle existieren. Ein pH-Gradient wird z.B. in der Nähe der Zellmembran durch die chemischen Reaktionen der Atmungskette verursacht, Gradienten von Nährstoffmolekülen entstehen, weil bei der Substrataufnahme durch die Zellmembran transportierte Moleküle hauptsächlich über Diffusionsvorgänge ins Zellinnere gelangen. Obwohl solche Gradienten mit partiellen DGL beschrieben werden können (Giersch 2003), wird bei den vorliegenden Modellen eine einheitliche Verteilung der Moleküle innerhalb der Zelle angenommen, da räumliche Inhomogenitäten experimentell nicht erfaßt werden können (s. Punkt 1 und Punkt 3).
5. Zellen passen sich an Veränderungen der Umgebung an, indem die Enzymkonzentrationen in der Zelle durch Induktion der betreffenden Gene variiert werden. Mehrere Modelle, welche Vorgänge der Transkription, Translation und des Metabolismus integrieren, sind erstellt worden (Kremling et al. 2001; Kremling 2001). Um die erheblich differierenden Zeitskalen, auf denen die verschiedenen Vorgänge stattfinden, zu berücksichtigen, benutzen die Modelle zeitverzögerte DGL, vgl. z.B. (Santillán und Mackey 2001). Innerhalb eines Beobachtungszeitraums von 20-40 s können jedoch durch den Substrat-Puls ausgelöste Vorgänge der genetischen Regulation mit einiger Sicherheit ausgeschlossen werden, s. Abbildung 2-2, und sind daher nicht in die vorliegenden Modelle integriert worden.
6. Für die exakte Beschreibung der Reaktionsgeschwindigkeit von Spezies, die in nur sehr geringen Mengen in der Zelle vorhanden sind ist die absolute Zahl der Moleküle von Bedeutung, weil z.B. die Bindung von Molekülen an die Komponenten oder deren Replikation in der Regel als stochastische Prozesse erachtet werden. Viele stochastische Modelle sind bekannt (Kuo und Keasling 1996; Morton-Firth und Bray 1998), wobei sich die meisten von ihnen mit den Vorgängen der Genregulation beschäftigen (Gibson und Bruck 2000; Kierzik 2002). Die im Rahmen dieser Arbeit erstellten Modelle beschreiben metabolische Vorgänge, d.h. wird von einem durchschnittlichen Zellvolumen von $0,5 \mu\text{m}^3$ (*E.coli* bei einer Verdünnungsrate von $0,125 \text{ h}^{-1}$ (Pramanik und Keasling 1997)) und einer durchschnittlichen Metabolitkonzentration von $1 \mu\text{M}$ ausgegangen, so ist die Gesamtzahl der Moleküle in der Zelle größer als $6 \cdot 10^8$. Daher kann ein deterministisches Modell erstellt werden.

Zusätzlich wurden im Rahmen dieser Arbeit noch weitere allgemeine Hypothesen angenommen, die im Folgenden besprochen werden.

Das Stoffwechselnetzwerk besteht aus zahllosen Reaktionen, die eng miteinander verwoben sind und durch etliche Effektoren beeinflusst werden. Hinzu kommt, daß einige Stoffwechselwege in Abhängigkeit von bisher unbekannten Bedingungen aktiv oder inaktiv sein können. Ein Modell, das den gesamten Metabolismus beschreibt, wäre also äußerst komplex und könnte bei dem derzeitigen Wissensstand nicht ohne weiteres erstellt werden (Gombert und Nielsen 2000). In der vorliegenden Arbeit wurde

^{3.8} Auch wenn eine räumliche Separierung bestimmter Spezies vorliegen würde, könnten diese aufgrund des Zellaufschlusses experimentell nicht getrennt voneinander quantifiziert werden.

die Wahl der abhängigen Variablen des DGL-Systems, d.h. der Metabolitkonzentrationen deren zeitliche Änderung simuliert werden sollte, im Wesentlichen davon beeinflusst, ob und in welcher Qualität experimentelle Daten für die jeweiligen Stoffwechselprodukte vorhanden waren. Dazu wurden die in Abschnitt 2.2.3 dargestellten Methoden für die Vereinfachung der Reaktionsnetzwerke verwendet, um z.B. nicht alle Metabolite eines linearen Reaktionswegs beschreiben zu müssen, oder Metabolite, deren Meßsignale nicht getrennt voneinander detektiert werden konnten^{3,9}, im Modell zusammenzuziehen. Auf diese Weise sollte vermieden werden, daß zu viele Stoffwechselprodukte mit unbekannten Anfangswerten^{3,10} in das DGL-System einfließen und aufgrund der resultierenden Unter-Identifizierbarkeit die Aussagequalität des Modells reduzieren, s. Abschnitt 3.4.2 (S. 41). Abhängige Variablen zur Beschreibung des zeitlichen Verlaufes der Nukleotide, wie z.B. ADP, ATP, NAD, wurden nicht eingesetzt und statt dessen die gemessenen Trajektorien als zeitabhängige Modellparameter verwendet. Hier galt die Annahme, daß die Nukleotide an sehr vielen biochemischen Reaktionen als Kofaktoren beteiligt sind, so daß ihre Konzentrationsänderungen vermutlich nicht ohne weiteres allein auf Basis der gemessenen Einflußgrößen erklärbar ist. Um zu prüfen, inwiefern die Wahl der abhängigen Variablen die Ergebnisse beeinflusst, wurden mehrere Modelle mit jeweils anderen Netzwerkstrukturen erstellt.

Als Terme für die Geschwindigkeiten der enzymkatalysierten Reaktionen wurden bekannte, *in-vitro* bestimmte Kinetiken der jeweiligen Enzyme eingesetzt. Die Informationen über die durch das Enzym katalysierten Reaktionen, die algebraische Gleichung der Kinetik, die bestimmten Parameterwerte und die Struktur des Enzyms sind über Online-Datenbanken, wie (Appel et al. 1994; Kanehisa und Goto 2000; Karp et al. 2000; Schomburg et al. 2002), zugänglich. Häufig werden in unterschiedlichen Datenbanken oder Literaturstellen widersprüchliche Angaben über die Art und Zahl der Effektoren, die kinetischen Gleichungen oder die Parameterwerte gemacht. Dies kann damit begründet werden, daß die mit dem gleichen Enzym durchgeführten Experimente oft von verschiedenen Versuchsbedingungen, wie z.B. unterschiedlichen Enzymaktivitäten, Effektoren oder physikalischen Bedingungen (pH-Wert, Temperatur, Salzkonzentration, etc.), ausgehen. Es läßt sich also z.B. auf der Basis von Plausibilitätstests der einzelnen Angaben nicht ohne weiteres entscheiden, welche Informationen zur Modellbildung genutzt werden können. Daher wurden bereits an dieser Stelle des Modellierungsprozesses viele konkurrierende Modelle zur Berücksichtigung sich gegenseitig ausschließender kinetischer Gleichungen erstellt, s. Kapitel 5.

Die in den Online-Datenbanken angegebenen Werte für einen Parameter können um einen Faktor von 10-1.000 variieren. Gleichzeitig ist zu vermuten, daß die die *in-vitro* bestimmten Parameterwerte unter *in-vivo* Bedingungen nicht mehr gültig sind, s. Abschnitt 2.2.3. Deshalb wurde ausgehend von den als Startwerten eingesetzten *in-vitro* Parameterwerten – soweit sie bekannt waren – über die iterative Veränderung der Werte eine optimale Konstellation bestimmt, für die eine Zielfunktion, als zeitintegrales Maß des Abweichungsgrades zwischen den simulierten und experimentellen Trajektorien, ein Minimum aufwies. D.h. ein Optimierungsproblem, die Minimierung der Zielfunktion, wurde durch das Fitten der Parameterwerte gelöst. Für die Optimierung erschwerend, s. Abschnitt 3.4, wirkt sich aus, daß die Parameteranzahl in den Kinetiken relativ groß ist. Eine Zweisubstrat-Kinetik mit einem Effektor hat z.B. 5-7 Parametern, s. Anhang E. Aus diesem Grund wurden in der Regel vereinfachte enzymkinetische Gleichungen, wie z.B. multiplikative Formen der Michaelis-Menten-Gleichung, s. Abschnitt 2.2.3, eingesetzt.

^{3,9}S. (Oldiges 2004) für die Zusammensetzung des zur Kalibrierung verwendeten externen Standards für das Substanz-Gemisch.

^{3,10}Die in (Chassagnole et al. 2002; Vaseghi et al. 1999) verwendete Methode, Anfangswerte experimentell nicht bestimmter Komponenten unter der Annahme, daß diese sich mit meßbaren Spezies nahezu im Fließgleichgewichtszustand befinden, abzuschätzen, wird aufgrund der hier vorliegenden stark transienten Bedingungen nicht herangezogen.

3.4 Computermodell

3.4.1 Erstellung und Verifikation des Computermodells

Die Implementierung des DGL-Systems erfolgte über die Eingabe der Gleichungen in die Oberfläche von MMT bzw. in eine SBML-Datei im Fall von MMT2 (SBML: Systems Biology Markup Language, eine XML-Spezifikation zur einheitlichen Beschreibung von im Gebiet der Systembiologie erstellten Modellen (Finney und Hucka 2003; Hucka et al. 2003))^{3.11}. Spezielle Tests zur Prüfung dieses Arbeitsschritts wurden nicht angewendet, da die Syntaxprüfung der Compiler die Mehrzahl der Fehler während der Übersetzung der Eingaben in ein ausführbares Computerprogramm abfängt. Trotzdem ist noch die fehlerhafte Angabe von Kinetiken bestimmter Reaktionen oder von Effektoren bzw. Parameterwerten in den Reaktionskinetiken möglich. Solche Fehler können im Allgemeinen ausgeschlossen werden, indem die Modelle vor der Optimierung mit den gewählten Startwerten simuliert und die Ergebnisse überprüft werden.

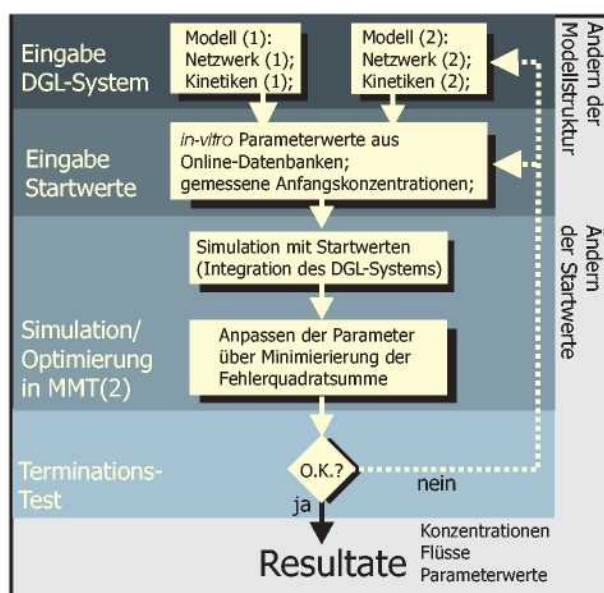


Abb. 3–3 Ablaufschema zur Erstellung des Computermodells

Nach Eingabe der Gleichungen in die Softwarewerkzeuge und der Übersetzung der Gleichungen in ein Computermodell, werden im Wesentlichen zwei automatisierte Rechnungen auf der Basis der Modelle durchgeführt: die Simulation, d.h. die numerische Integration des DGL-Systems und die Optimierung, s. auch Abb. 3–3. Dafür wurden die genannten Softwarewerkzeuge in dieser Arbeit zum ersten Mal in größerem Umfang verwendet. Daneben wird z.B. in den Lehrbüchern zur VV&T immer darauf hingewiesen, daß auch automatische Operationen stets ein gewisses Fehlerrisiko in sich bergen, auch wenn sie von kommerziellen, häufig benutzten Softwarewerkzeugen ausgeführt werden^{3.12}. Daher wurden die erhaltenen Simulationsergebnisse im Rahmen dieser Arbeit abgesichert, indem die hauptsächlich verwendete Software MMT zunächst dahingehend getestet wurde, ob die implementierten Methoden für die Problemstellungen der Simulation und Optimierung generell geeignet waren, s. unten. Weiterhin

^{3.11}Genauer gesagt handelt es sich um eine Spezifikation von SBML Level 1 (Hucka et al. 2001), da für die richtige Wiedergabe der vorliegenden Modelle noch einige zusätzliche Komponenten z.B. zur Beschreibung der chemischen Reaktionskinetiken oder der experimentellen Daten eingeführt wurden.

^{3.12}S. z.B. (Page 1991): „In der Praxis kommt es nicht selten vor, daß ein Simulationsprogramm völlig andere Werte liefert, als man aufgrund des konzeptuellen Modells erwarten würde. Problematischer sind jedoch Fehler, die nicht auf den ersten Blick zu erkennen sind. Auch plausibel erscheinende Ergebnisse sollten daher in systematisch ausgewählten Fällen nachgerechnet werden, um mögliche Programmfehler aufzudecken.“

wurde über die Prüfung des Quelltextes und Black-Box-Tests sichergestellt, daß die Methoden korrekt implementiert und ausgeführt wurden. Dabei konnten kleinere Fehler behoben werden. Die Prüfung der während der Testphase von MMT entstandenen Software MMT2 war weniger aufwendig, da existierende Modelle von MMT auf MMT2 übertragen und die Rechenergebnisse beider Werkzeuge abgeglichen werden konnten.

Einige Eigenschaften der Modelle zur Auswertung von Substrat-Puls-Experimenten stellen besondere Anforderungen an die Simulation und Optimierung, die bereits bei der Programmierung der Softwarewerkzeuge durch Implementierung spezieller Routinen berücksichtigt wurden. Im Folgenden wird eine Übersicht über die als feste Bestandteile von MMT bzw. MMT2 ohne weitere Änderungen übernommenen Routinen gegeben und ihre Eignung zur Lösung der jeweiligen Aufgaben diskutiert, s. auch Anhang F für detailliertere Ausführungen.

Bei der numerischen Integration eines DGL-Systems ist immer die Genauigkeit, bzw. Konvergenz, und die Stabilität der Lösung von großer Bedeutung, s. Anhang F.1. Zusätzlich kann hier aufgrund der Verwendung chemischer Kinetiken und der Beschreibung einer sich zu einem Zeitpunkt schlagartig ändernden Größe einer Variablen (Glucose-Puls) ein steifes DGL-System resultieren, wie es bei Systemen, die auf sehr unterschiedlichen Zeitskalen operieren, häufig vorkommt (Aiken 1985). Unter praktischen Aspekten lassen sich steife Systeme als DGL definieren, zu deren Lösung implizite Integrationsverfahren erheblich besser geeignet sind als explizite (Curtiss und Hirschfelder 1952). Während in diesem Fall Stabilitätsanforderungen die Integrationsschrittweite von expliziten Verfahren bestimmen, können implizite Algorithmen wesentlich größere Schrittweiten bei gleicher Genauigkeit der Lösung verwenden, wodurch der Zeitaufwand bei der Integration verringert wird. Obwohl es in der Literatur keine einheitliche Definition der Steifheit gibt, lassen sich einige charakteristische Eigenschaften steifer Systeme herausstellen. Sie enthalten Lösungskomponenten, die sich nur langsam ändern. Störungen dieser langsam veränderlichen Lösungen werden jedoch sehr schnell gedämpft. Äußere Kennzeichen sind bei Algorithmen mit implementierter Schrittweitensteuerung eine relative Unabhängigkeit der Schrittzahl von der vorgegebenen Genauigkeit und eine große Anzahl wiederholter Schritte. Weitere Hinweise können nicht glatte Lösungsfunktionen liefern oder aber durch die Analyse der Jacobi-Matrix^{3.13} linearer Systeme gefunden werden (Hairer und Wanner 1991). Die in MMT und MMT2 implementierte, frei erhältliche Routine zur numerischen Integration des DGL-Systems, LSODA (Petzold 1983; Petzold und Hindmarsh 2004), ist eine Variante des DGL-Lösers, LSODE (Livermore Solver for Ordinary Differential Equations (Hindmarsh 1980)). Sie verwendet variable Integrationsschrittweiten und beinhaltet Lösungsalgorithmen für steife und nicht-steife DGL-Systeme, zwischen denen automatisch gewechselt wird. Damit wird die numerische Integration der vorliegenden Modelle durch LSODA als hinreichend genau betrachtet.

Wie bereits im letzten Abschnitt erwähnt, werden die genauen Parameterwerte ermittelt, indem der Wert einer Zielfunktion, die die zeitintegrale Abweichung zwischen Meßdaten und simulierten Trajektorien wiedergibt, minimiert wird. Die in MMT bzw. MMT2 implementierte Zielfunktion läßt sich mathematisch über die folgende Gleichung ausdrücken:

$$F(\mathbf{p}, \mathbf{h}, \mathbf{c}_0, \tilde{\mathbf{c}}, m) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m h_j (c_{ij}(\mathbf{p}, \mathbf{c}, \mathbf{c}_0) - \tilde{c}_{ij})^2, \quad (3-2)$$

mit $\mathbf{p}$ dem Vektor mit den aktuellen Parameterwerten, $\mathbf{c}_0$ dem Vektor mit den Anfangskonzentrationen, $i = 1 \dots n$ bzw. $j = 1 \dots m$ den Indizes für die Zeitpunkte bzw. für die verschiedenen Metabolite, $\mathbf{c}$

^{3.13} Alle Eigenwerte der Jacobi-Matrix linearer Systeme besitzen negative Realteile und der Steifigkeitsquotient ist groß; für $\dot{\mathbf{y}} = \mathbf{A}\mathbf{y} + \mathbf{b}(t)$ heißt: $\max_{\lambda \in \Lambda(A)} |\operatorname{Re} \lambda| / \min_{\lambda \in \Lambda(A)} |\operatorname{Re} \lambda|$, Steifigkeitskoeffizient; λ bezeichnet einen Eigenwert, $\Lambda(A)$ die Diagonalmatrix der Eigenwerte; s. Anhang F für die Definition der Jacobi-Matrix.

bzw. $\tilde{c}$ dem Vektor der simulierten Trajektorien bzw. dem Vektor der experimentellen Trajektorien, $\tilde{c}_{ij}$ bzw. $c_{ij}(\mathbf{y}_i, \mathbf{p})$ einem Meßdatenpunkt bzw. dem simulierten Wert des Metaboliten j zum Zeitpunkt i und mit h_j einem von dem Metaboliten j abhängigen Wichtungsfaktor für die quadrierten Residuen. Die Zielfunktion ist also eine gewichtete Fehlerquadratsumme, wobei nur in MMT Wichtungsfaktoren für einzelne Metabolite angegeben werden können, während bei Verwendung von MMT2^{3,14} gilt: $h_j = 1 \forall j$. Neben Gleichung (3–2) werden häufig auch andere Zielfunktionen verwendet, wie hauptsächlich die Likelihood-Funktion, die die Abweichungen zwischen Modell und Daten noch mit weiteren Größen, wie z.B. mit den Varianzen der Meßdaten, verrechnet. Da bei ihrer Anwendung eine heteroskedastische Verteilung der Varianzen berücksichtigt werden kann, d.h. die Varianzen der abhängigen Größen (Konzentrationen) sind eine Funktion der unabhängigen Größe (Zeit), wäre sie vermutlich als Zielfunktion besser geeignet. Dennoch wird die Fehlerquadratsumme zunächst als ausreichend angesehen, weil die benötigten Informationen zur Berechnung der Likelihood-Funktion aus den vorliegenden Experimenten nicht ohne weiteres mit hinreichender Genauigkeit erhältlich sind^{3,15}.

Die Optimierungsaufgabe lautet mit der vereinfacht dargestellten Gleichung (3–2) formal:

$$\hat{F} = \min \{F(\mathbf{p}) \mid \mathbf{p} \in \mathcal{R}\} \Leftrightarrow \arg \min(F, \mathcal{R}) = \min \left\{ \mathbf{p} \in \mathcal{R} \mid F(\mathbf{p}) = \hat{F} \right\}. \quad (3-3)$$

D.h. ein Minimum der Zielfunktion, $\hat{F}$, wird über der zulässigen Menge, $\mathcal{R} = \{\mathbf{p} \in \mathbb{R}^q : \mathbf{p} \geq \mathbf{0}\}$, gesucht, wobei die Menge der Punkte, die angenommen werden können, mit $\arg \min(F, \mathcal{R})$ bezeichnet wird. Da Gleichung (3–3) aufgrund der Größe des Parameterraums, $\mathbb{R}^q$, analytisch nicht gelöst werden kann, wurde ein numerisches Verfahren, die Subplex-Methode (Rowan 1990), zur iterativen Lösung der Aufgabe implementiert. Sie erweitert die Nelder-Mead Simplex-Methode (Nelder und Mead 1965), die als direktes Suchverfahren, d.h. Gradienten-freie Methode, zur Optimierung der hier vorliegenden, ver-räuschten Zielfunktionen besonders geeignet ist. Der Speicherbedarf der Nelder-Mead-Methode steigt jedoch quadratisch zur Dimension q des Parametervektors an und die benötigte Rechenzeit für einen Iterationsschritt nimmt ebenfalls für größere q stark zu, weshalb der Subplex-Algorithmus die Effizienz des Verfahrens verbessern soll. Dazu wird der Parameterraum in Subräume zerlegt, in denen der Nelder-Mead-Algorithmus das Minimum der Zielfunktion bzgl. jeweils einiger Komponenten des Parametervektors bestimmt, wobei die anderen Komponenten ihren Wert aus dem vorherigen Iterationsschritt beibehalten. So werden die positiven Eigenschaften des Nelder-Mead-Verfahrens, insbesondere die Robustheit des Algorithmus bzgl. der Streuung der Meßpunkte, beibehalten, wobei der Speicherbedarf nunmehr linear von der Dimension des Problems abhängig ist, und die Rechenzeit sich erfahrungsgemäß stark verkürzt (Rowan 1990).

Neben dem Subplex-Algorithmus gibt es eine Vielzahl anderer Optimierungsverfahren, s. (Press et al. 1992) bzw. Anhang F.2 für eine grobe Klassifizierung und den Vergleich mit der dort genauer beschriebenen Subplex- bzw. Nelder-Mead-Methode. Indirekte Such- oder Gradienten-Verfahren, die die erste oder zweite Ableitung der Zielfunktion zur Optimierung bestimmen, erfordern für das gegebene hoch-dimensionale Problem einen hohen Rechenaufwand pro Iterationsschritt. Zusätzlich ist gezeigt worden (Hurlebaus 2001), daß die ableitungsfreien, direkten Suchverfahren in Anbetracht der vorliegenden unstetigen Zielfunktion vorzuziehen sind. Ob andere bekannte, direkte Suchverfahren bessere Eigenschaften haben als der Subplex-Algorithmus, ist nicht ohne weiteres ersichtlich: die zyklische Koordinatenmethode, s. (Fletcher 1987), stellt z.B. nur einen Spezialfall des Subplex-Algorithmus dar, und, wie andere Algorithmen die Suchrichtung in jedem Iterationsschritt optimieren (Hooke und Jeeves 1961; Rosenbrock 1960), verbessert auch der Subplex-Algorithmus in jedem Schritt die Aufteilung der

^{3,14}Im Rahmen dieser Arbeit wurden MMT2-Versionen bis Version 3.3-5 verwendet.

^{3,15}Im Zuge der operationalen Modellvalidierung werden noch andere Kriterien zur weitergehenden Beurteilung der Regressionsgüte ausgewählt, bereits auf der Basis von Gleichung (3–2) angepaßter Modelle der Fehlerquadratsumme gegenübergestellt, s. Abschnitt 6.3. Diese sind vermutlich bei einer verbesserten Ausgangssituation als Zielfunktionen zur Parameteranpassung vorzuziehen.

Subräume. Allerdings gehört der Subplex-Algorithmus zu den Optimierungsverfahren, die, weil sie für die Suche nach lokalen Extrema entwickelt wurden, auch als „lokal“ bezeichnet werden, s. (Aarts und Lenstra 1997). Im Gegensatz dazu können die sogenannten „globalen“ Suchverfahren (Rinnooy Kan und Timmer 1989) durch die Hinzunahme von stochastischen Elementen, wie z.B. durch die zufällige Änderung des Parametervektors in jedem Iterationsschritt bei den Stochastischen Algorithmen, das globale Optimum mit höherer Sicherheit bestimmen. Die Effizienz solcher Verfahren, zu denen u.a. auch die Genetischen Algorithmen gehören, müßte jedoch in Anbetracht der hohen Parameterzahl der Modelle ($50 \leq q \leq 150$) erst noch getestet werden. Zur Verbesserung des Optimierungsergebnisses auf Basis des nur lokalen Subplex-Suchverfahrens konnten mit MMT jedoch zusätzlich noch zwei Strategien verwendet werden. Zum einen war die Einschränkung des Suchraums über die Vorgabe von Grenzen für die Parameterwerte, möglich. Zum anderen war in MMT eine adaptierte „Random- Restart-Hill-Climbing“-Methode (Russell und Norvig 1995, S. 111-115) implementiert worden (Hurlebaus 2001). Dies bedeutet, daß der optimierte Parametervektor nach Beendigung des Subplex-Verfahrens innerhalb gewisser Bereiche zufällig geändert, der Optimierungsalgorithmus dann automatisch erneut gestartet, und dadurch insgesamt ein größerer Raum abgesucht werden konnte.

Dennoch traten in einigen Fällen, insbesondere dann, wenn Rohdaten ohne vorherige Glättung oder Filterung zur Berechnung der Zielfunktion verwendet wurden, aber auch, wenn eine hohe Zahl an Parametern angepaßt werden sollte, Schwierigkeiten bzgl. der Konvergenz des Verfahrens auf. Evtl. weist dies doch darauf hin, daß noch andere Optimierungsalgorithmen getestet werden sollten. Der Multiple-Shooting-Algorithmus (van Domselaar und Hemker 1975; Bock 1981; Bock 1983) besitzt z.B. gute Konvergenzeigenschaften für das dargestellte Problem (Bock 1981) und wird bereits in (bio)-technologischen Bereichen zu ähnlichen Zwecken verwendet (Timmer et al. 1998; Oberle und Sothman 1999; Horbelt et al. 2000; Horbelt et al. 2002; Müller et al. 2002). Möglicherweise könnte auch eine Kombination von globalen Suchverfahren für eine erste grobe Suche des Optimums und lokalen direkten Verfahren für eine verfeinerte Suche den Optimierungsvorgang verbessern. In (Bock 1983) wurden allerdings Beispiele demonstriert, für die –völlig unabhängig von der gewählten Optimierungsstrategie– kein Minimum gefunden werden konnte, so daß im Rahmen dieser Arbeit auf vertiefte Untersuchungen der verschiedenen Verfahren verzichtet und statt dessen Veränderungen an den Modellen vorgenommen wurden, falls kein Minimum identifiziert werden konnte.

Neben MMT und MMT2 wurde durch andere Gruppen eine Reihe weiterer Softwarewerkzeuge zur Simulation und Optimierung nichtlinearer DGL-Modelle für die Beschreibung von Zellvorgängen entwickelt. Diese implementieren teilweise andere Integrations- und Optimierungsalgorithmen, s. (Arkin 2001a) für einen Überblick. Solche für spezielle Fragestellungen entwickelte Softwarewerkzeuge haben im Gegensatz zu universell einsetzbaren Simulationssprachen, wie Maple oder Matlab, immer den Vorteil, daß die Möglichkeit der Fehler bei der Modellierung und Simulation reduziert wird, andererseits ist aber auch die Flexibilität der Modellbildung verringert, s. Abschnitt 2.3.2. Daher wurden basierend auf den gesammelten Erfahrungen Anforderungen definiert, die ein Softwarewerkzeug zur Auswertung von Substrat-Puls-Experimenten im Wesentlichen erfüllen sollte, und andere, gängige Softwarewerkzeuge hinsichtlich der diesbezüglich gesteigerten Möglichkeiten bzw. Einschränkungen mit MMT und MMT2 verglichen. Dabei zeigte sich, daß die frei erhältliche Software Gepasi (künftig: Copasi) (Mendes und Kell 1998) und MMT bzw. MMT2 am besten für die Erstellung und Parameteranpassung großer DGL-Systeme geeignet sind.

3.4.2 Operationale Modellvalidierung

Im letzten Schritt der Modellbildung und Validierung werden die Ausgaben des Computermodells mit dem Verhalten des realen Systems verglichen und das Modell auf seine Struktur- und Verhaltensgültigkeit überprüft. Für einen objektiven Vergleich werden neben den Meßdaten, die bereits zur Parameteranpassung der Modelle verwendet wurden, zusätzliche, unabhängig erhobene Datensätze benötigt (Page 1991). Solche Daten waren im vorliegenden Fall nicht erhältlich. Daher wurde für die notwendige weiterführende Validierung entschieden:

1. die Qualität der Modellaussagen zunächst auf Basis der vorhandenen Meß- und Literaturdaten zu prüfen. Aufgrund der großen Freiheit bei der Wahl des Reaktionsnetzwerkes, der Kinetiken und der Parameterwerte, s. Abschnitt 3.3, sollten diesen elementaren Kontrollschritt mehrere Modelle überstehen (Nelles 2001);
2. die Anpassungsgüte der verbleibenden Modelle hinsichtlich anderer als der bereits unter Punkt 1 verwendeten Kriterien zu untersuchen, um die Zahl der Modelle zu reduzieren;
3. die Modelle selbst, ohne Hinzunahme von Meßdaten, eingehender zu prüfen, um die Modellfamilie weiter einzugrenzen;

Bei der beschriebenen Vorgehensweise zur Auswertung von Substrat-Puls-Experimenten lassen sich die folgenden Modellausgaben erwarten: die optimierten Parameterwerte, die im Parameteroptimum simulierten Trajektorien der intrazellulären Metabolitkonzentrationen bzw. der Reaktionsgeschwindigkeiten (Flüsse) und die minimierte Fehlerquadratsumme. Alle Ausgaben wurden weiterführenden vergleichenden Analysen unterworfen, s. Punkt 1.

Aufgrund der Schwierigkeiten beim Auffinden eines globalen Minimums der Fehlerfunktion, s. letzten Abschnitt, wurde die Fehlerquadratsumme nur in Kombination mit graphischen Vergleichen herangezogen, um die Qualität der Wiedergabe der experimentellen Daten durch die simulierten Konzentrationsverläufe zu beurteilen. Da in WT1 vor dem Puls ein Fließgleichgewicht aufrechterhalten wurde, ergab sich eine zusätzliche Bedingung für die Flußverteilung, d.h. die Werte der simulierten Flüsse sollten innerhalb dieses Zeitraums nahezu konstant sein. Sie können im stationären Zustand mit veröffentlichten Ergebnissen von unabhängigen, unter ähnlichen Bedingungen durchgeführten metabolischen Flußanalysen, s. Abschnitt 2.2.2, verglichen werden, während für den Vergleich charakteristischer Flußverläufe im transienten Zustand bereits veröffentlichte Simulationen, vorwiegend (Sauter et al. 2002; Chassagnole et al. 2002), zur Verfügung stehen. Nur solche Modelle, die diese Prüfungen weitestgehend bestanden, wurden noch vertieft im Sinne der operationalen Validierung untersucht, während ansonsten die Modellstrukturen oder die Gleichungen der Reaktionskinetiken aufgrund der gewonnenen Erkenntnisse abgeändert wurden. Daher werden die graphischen Vergleiche bereits in Kapitel 5 behandelt, um den iterativ unter Rückkopplung ablaufenden Prozeß der Modellerstellung besser verdeutlichen zu können.

Zur verbesserten Untersuchung der Anpassungsgüte, s. Punkt 2, wurden gängige statistische Verfahren eingesetzt, die bisher vorwiegend bei Regressionsmodellen Verwendung fanden. Oft benutzte, statistische, skalare Kennzahlen für den Grad der Abweichung zwischen den Daten und dem (Regressions-)Modell sind der Wert der Likelihood-Funktion und das Akaike-Kriterium, s. z.B. (Burnham und Anderson 1998). Bei ihrer Berechnung werden neben den Residuen, die den absoluten Wert der Differenz zwischen Modell- und der Meßgröße zu einem Zeitpunkt wiedergeben, noch weitere Informationen berücksichtigt, wie z.B. die Varianzen der Residuen oder die Zahl der Modellparameter, s. Abschnitt 6.3. Häufig werden auch Hypothesentests, s. Abschnitt 2.3.2 (S. 22) verwendet, um ein Modell bzgl. seiner Anpassungsgüte zu akzeptieren oder abzulehnen. Besonders geeignet erschien hier ein Hypothesentest,

der auf der mehrfachen Anpassung eines Regressionsmodells an zahlreiche, mit einer *Wild Bootstrap*-Näherung zufällig erzeugte Datensätze beruht (Stute 1997; Stute et al. 1998). Denn für das Gelingen des Tests werden keinerlei Forderungen an die statistische Verteilung der Meßunsicherheiten gestellt. Darüber hinaus wurde in (Stute et al. 1998) anhand von kleineren, nichtlinearen Regressionsmodellen und auf ihnen beruhenden *in-silico* Datensätzen^{3.16} gezeigt, daß ein Regressionsmodell mit mehr Parametern als das dem künstlichen Datensatz zugrundeliegende Modell von dem Hypothesentest korrekt abgelehnt werden konnte. Dies demonstriert eindrucklich das Potential der Methode. Aus diesem Grund wurde das Verfahren im Rahmen dieser Arbeit für die Beurteilung der Anpassungsgüte von multivariaten, nichtlinearen DGL-Modellen erweitert und seine Verwendbarkeit anhand von beispielhaften Modellen untersucht, s. Abschnitt 6.3.

Wenn es, wie in diesem Fall erwartet, eine begrenzte Anzahl von Modellen mit unterschiedlicher Struktur aber ähnlicher Anpassungsgüte gibt, bleiben die Fragen offen, ob (a) auf der Basis der Meßdaten eine beste Modellstruktur ausgewählt, und (b) bei der Parameterschätzung eine eindeutige Lösung des korrespondierenden Parametersatzes gefunden werden kann, s. Punkt 3. Für den Idealfall, bei dem der Informationsgehalt der Meßdaten nicht durch ein überlagertes Rauschen reduziert wird^{3.17}, beschäftigen sich Studien zur Unterscheidbarkeit (bzgl. a) und Identifizierbarkeit von Modellen (bzgl. b) mit diesen Fragen, s. (Walter 1982).

Definition 3.4.1 Eine Modellstruktur $\hat{M}(\cdot)$ ist anhand ihrer Ausgaben strukturell von einer Modellstruktur $M(\cdot)$ unterscheidbar, wenn für nahezu alle Parametervektoren $\mathbf{p} \in \mathcal{P}$ die Gleichung:

$$\begin{cases} \hat{\mathbf{p}} \in \hat{\mathcal{P}}, \\ \hat{\mathbf{y}}(t, \hat{\mathbf{p}}, \mathbf{u}) = \mathbf{y}(t, \mathbf{p}, \mathbf{u}), \quad \forall t \in \mathbb{R}^+, \forall \mathbf{u}(\cdot) \in \mathcal{U} \end{cases}$$

keine Lösung für $\hat{\mathbf{p}}$ besitzt, wobei $\mathbf{u}(\cdot) \in \mathcal{U}$ die Eingangsgrößen des Modell, $\mathbf{y}$ bzw. $\hat{\mathbf{y}}$ die Modellausgaben und $\mathcal{P}$ bzw. $\hat{\mathcal{P}}$ jeweils zulässige Lösungsräume für die beiden Parametervektoren darstellen (Walter et al. 1984).

Ein Beweis der Unterscheidbarkeit ist aufgrund der Größe der vorliegenden Modelle mit den bisher beschriebenen Methoden (Walter et al. 1985) nicht durchführbar, daher wurde zur weiteren Prüfung der Modelle nur die Identifizierbarkeit untersucht.

Definition 3.4.2 $M_i(\mathbf{p}_i) \equiv M_j(\mathbf{p}_j)$ soll bedeuten, daß das Modell mit der Struktur $M_i(\cdot)$ und dem Parametervektor $\mathbf{p}_i$ für alle Eingaben und zu jedem Zeitpunkt das gleiche Eingabe-Ausgabe-Verhalten hat, wie das Modell mit der Struktur $M_j(\cdot)$ und dem Parametervektor $\mathbf{p}_j$. $M(\cdot)$ wird dann als eindeutig identifizierbar (oder global identifizierbar) bezeichnet, wenn für nahezu alle $\mathbf{p}^*$ die einzige Lösung $\mathbf{p}$, für die $M(\mathbf{p}) \equiv M(\mathbf{p}^*)$ gilt, $\mathbf{p} = \mathbf{p}^*$ ist (Walter und Pronzato 1996).

Mit anderen Worten gibt es für eine nicht eindeutig identifizierbare Modellstruktur $M(\cdot)$ mehrere Parametervektoren, für die das Eingabe-Ausgabe-Verhalten des Modells identisch ist und von denen keine Vektoren unter Zuhilfenahme der Meßdaten eliminiert werden können. Viele Methoden zur Überprüfung der theoretischen bzw. strukturellen Identifizierbarkeit, wurden entwickelt, s. (Walter 1982; Chappell und Godfrey 1990; Ljung und Glad 1994b; Walter und Pronzato 1996), und sind u.a. auch im Bereich der Biotechnologie oder technischen Chemie auf nichtlineare DGL-Modelle angewendet worden, s. z.B. (Julien et al. 1998; Müller et al. 2002). Für größere Modelle ist jedoch der Nachweis der theoretischen Identifizierbarkeit oft nicht ohne weiteres möglich, da bei der Anwendung der Methoden

^{3.16}Unter *in-silico* Datensätzen werden künstlich, mit dem Rechner aus Modellen gebildete Meßdaten verstanden.

^{3.17}Hierzu werden in den meisten Fällen die Daten *in-silico* aus einer der zu untersuchenden Modellstrukturen und dem zugehörigen Parametervektor generiert.

komplizierte mathematische Ausdrücke resultieren, deren Lösung äußerst rechenintensiv ist. Gleichzeitig konnte gezeigt werden (Nihtilä und Virkkunen 1977), daß bereits die Parameter einer Monod-Kinetik^{3.18} in einem sehr einfachen DGL-Modell zwar theoretisch durch ein Batch-Experiment identifizierbar sind, aber in der Praxis aufgrund der verrauschten Meßdaten aus einem solchen Experiment nicht eindeutig bestimmt werden können (empirische Unter-Identifizierbarkeit des Modells). Damit wurde für die vorliegende Arbeit aus eher pragmatischen Gründen, d.h. in Anbetracht der Komplexität der vorliegenden Modelle und der Datenqualität, die Untersuchung der theoretischen Identifizierbarkeit der Modelle nicht in Betracht gezogen. Darüber hinaus wurde im vorliegenden Fall als wahrscheinlich angenommen, daß die Parameter der an dieser Stelle des Modellierungsprozesses existierenden Modelle auch praktisch nicht identifizierbar sind. Dies läßt sich neben der Datenqualität vorwiegend damit begründen, daß während der Erstellung der Modelle „*in-vitro* Kinetiken“ mit relativ vielen Parametern für die Beschreibung der Reaktionsgeschwindigkeiten eingesetzt werden, wobei alle Parameter geschätzt werden müssen. Denn die korrekten *in-vivo* Werte sind nicht bekannt und können somit bei der Optimierung nicht festgesetzt werden. Bevor also ein Vergleich der Modellstrukturen möglich ist, erscheint es zunächst notwendig, die vorliegenden Modelle auf die signifikanten Parameter zu reduzieren, d.h. die Parameter zu eliminieren, die über die gesamte Zeit hinweg, die Trajektorien aller Metabolitkonzentrationen nur gering beeinflussen. Dazu wurden innerhalb dieser Arbeit Methoden der Sensitivitätsanalyse (Saltelli et al. 2001) verwendet und weiterentwickelt (Jolliffe 1972; Jolliffe 1972; Turányi 1990; Froemel 2003), s. Abschnitt 6.2.4. Nach der Reduktion der Modelle ergeben sich evtl. inkonsistente Modellstrukturen, wodurch die Modellfamilie weiter eingegrenzt werden kann.

Dennoch wird durch die Modellreduktion noch nicht geklärt, ob die signifikanten Parameter auch eindeutig identifizierbar sind. Die Parameter der Kinetiken sind nämlich u.U. stark korreliert, so daß die Änderung eines Parameterwertes durch eine entsprechende Änderung eines anderen Parameterwertes kompensiert werden kann, wie Holmberg (1982) anhand einer graphischen Auftragung der Sensitivitätsfunktionen für das obige Beispiel der Monod-Kinetik demonstrierte^{3.19}. Für einfachere Modelle ist es möglich, eine Korrelation zwischen verschiedenen Parametern über die graphische Bestimmung ähnlich verlaufender Sensitivitätsfunktionen, aber auch über Monte-Carlo-Methoden relativ einfach nachzuweisen. Bei Monte-Carlo-Methoden werden die zu untersuchenden Parameter ausreichend häufig innerhalb gewisser Bereiche um den optimierten Wert zufällig geändert, und das Modell erneut simuliert bzw. optimiert s. z.B. (Press et al. 1992; Paker 1997; Saltelli et al. 2001; Hurlebaus 2001). Zur Ermittlung der Korrelationen werden dann bei Modellen mit wenigen Parametern die berechneten Werte der Zielfunktionen über die zugehörigen Parameterwerte aufgetragen. Für die Korrelationsanalyse von Modellen mit einer höheren Dimension eignet sich eher die Hauptkomponentenanalyse der aus den Sensitivitätsfunktionen gebildeten Kovarianzmatrix (Vajda et al. 1985), die auch in dieser Arbeit durchgeführt wurde, s. Abschnitt 6.2.5. Denn mit dieser Methode werden Linearkombinationen von Parametern ermittelt, die mit unterschiedlich starkem Ausmaß an den Simulationsergebnissen beteiligt sind. Die korrelierenden Parameter können dann über Vereinfachungen der Kinetiken zusammengefaßt werden, was zu einer weiteren Reduktion der Modelle führt, und evtl. den Ausschluß nicht mehr konsistenter Modelle möglich macht. Solche Strukturvereinfachungen sind jedoch bereits für die Parameter einer nichtlinearen Kinetik nicht ohne weiteres möglich, s. die dargestellten Gleichungen in Anhang D, und werden noch weniger bei Parametern aus Kinetiken, die verschiedenen Reaktionen zugeordnet wurden, gelingen. Im Rahmen dieser Arbeit werden daher solche Vereinfachungen nur beispielhaft durchgeführt, s. Abschnitt 6.2.5. Es muß damit für den in der vorliegenden Arbeit gewählten Ansatz zur Modellerstellung hingenommen werden, daß keine eindeutige Lösung für den Parametervektor gefunden werden kann und die verbleibenden Modellstrukturen nicht identifizierbar sind. Dies bedeutet, daß

^{3.18}Die Monod-Kinetik (Monod 1942) besitzt denselben mathematischen Ausdruck wie die Michaelis-Menten-Kinetik, sie wurde jedoch zur Beschreibung des Wachstums von Bakterienzellen abgeleitet.

^{3.19}Dies kann im übrigen auch für die hier vorliegenden Modelle gezeigt werden, s. Abb. 6–4 auf S. 101.

am Ende des beschriebenen Modellbildungs- und Validierungszyklus mehrere Modelle übrigbleiben, von denen keines mehr auf der Basis der Meßdaten ausgeschlossen werden kann.

Eine weiterführende Bewertung der Modelle, aufgrund der jedoch keine Modellstruktur verworfen werden sollte, gelingt über die Berechnung der Inversen der Fisher-Informationsmatrix (FIM), s. (Ljung und Glad 1994a), die unter Berücksichtigung der Meßunsicherheiten die beste erreichbare Schätzgenauigkeit der einzelnen Parameter in den verschiedenen Modellen angibt. Die Parameterschätzgenauigkeiten wurden somit auch in dieser Arbeit bestimmt, s. Abschnitt 6.2.6.

Zur sicheren Unterscheidung der Modelle sollten jedoch an dieser Stelle weitere Experimente durchgeführt werden. Dabei wäre ein sinnvoller Ansatz, Techniken der Versuchsplanung einzusetzen, s. z.B. (Takors 1997) für einen Überblick, d.h. über die Minimierung oder Maximierung verschiedener aus der FIM abgeleiteter Kriterien (Munack und Posten 1989; Baltes et al. 1994; Versyck et al. 1997), für jedes Modell optimale Versuchsbedingungen vorauszusagen, für die die Parameter praktisch identifiziert werden können.

Es bleibt jedoch die Frage offen, ob ein einmal auf diese Weise bestimmtes Modell das biologische System auch korrekt beschreibt, da im Kontext der Systembiologie häufig die Robustheit solcher Systeme diskutiert wird. Nach (Kitano 2002b) können die Eigenschaften robuster Systeme wie folgt klassifiziert werden:

1. Adaption an die Umgebung: bestimmte Merkmale der Systeme hängen nicht von Teilmengen der Umgebungsparameter ab;
2. Insensitivität von Parametern: nur eine schwache Reaktion des Systems erfolgt auf die Änderungen einiger kinetischer Parameter;
3. „Graceful Degradation“: langsames Drosseln der Systemfunktionen nach einer Beschädigung anstatt eines katastrophalen, sofortigen Ausfalls;

Sollen die DGL-Modelle also die robusten Eigenschaften des Systems wiedergeben, werden andere Kriterien zur Modellselektion und Versuchsplanung gefordert als sie im Rahmen dieser Arbeit zur operationalen Modellvalidierung vorgeschlagen werden. Die in der Einleitung des Kapitels definierte Fragestellung zielt jedoch darauf ab, Modelle zu finden, deren Ergebnisse für das Metabolic Engineering genutzt werden können. In dieser Hinsicht erscheint die gewählte Strategie sinnvoller, die jeweils signifikanten Parameter der Modelle zu ermitteln, deren Änderung eine Verbesserung der Stoffwechselleistung bewirken könnte, anstatt mit Hilfe von mathematisch komplizierten Kriterien insensitive, aber biologisch relevante Parameter zu identifizieren.

Nachdem nun die im Rahmen dieser Arbeit getroffene Wahl der Methoden zur Modellerstellung bzw. Validierung begründet wurden, und somit das erarbeitete Konzept der Dissertation umrissen ist, werden in den folgenden Kapiteln die wesentlichen Resultate der vorliegenden Arbeit dargestellt und diskutiert.

4 Validierung der Daten

4.1 Allgemeine Bemerkungen

Da die Validität der Daten eine notwendige Voraussetzung für den gesamten Modellbildungs- und Validierungsprozeß bildet, s. Abschnitt 2.3.2, sind die experimentellen Datensätze vor Beginn der Modellierung zu überprüfen. Aus diesem Grund wurde im Rahmen dieser Arbeit die Qualität der Datensätze über eine Fehlerrechnung, s. Abschnitt 4.2, und die Ermittlung der Nachweis-, Erfassungs- und Bestimmungsgrenzen der LC-MS-Methode beurteilt, s. Abschnitt 4.3. Auf Basis dieser bewährten Kriterien kann z.B. eine objektive Entscheidung darüber getroffen werden, welche Daten in das Modell eingehen und wie exakt diese durch das Modell beschrieben werden sollen. In Anbetracht der großen Datenmenge und der aufwendigen Prozedur der Substrat-Puls-Experimente, ist es wichtig, neben der Qualität auch die Konsistenz des Datensatzes zu prüfen. Dazu wurden hier Plausibilitätstests der Daten, s. Abschnitt 4.6, und eine statistische, induktive Methode, die Zeitreihenanalyse eingesetzt, s. Abschnitt 4.7.

4.2 Fehlerrechnung

4.2.1 Allgemeines

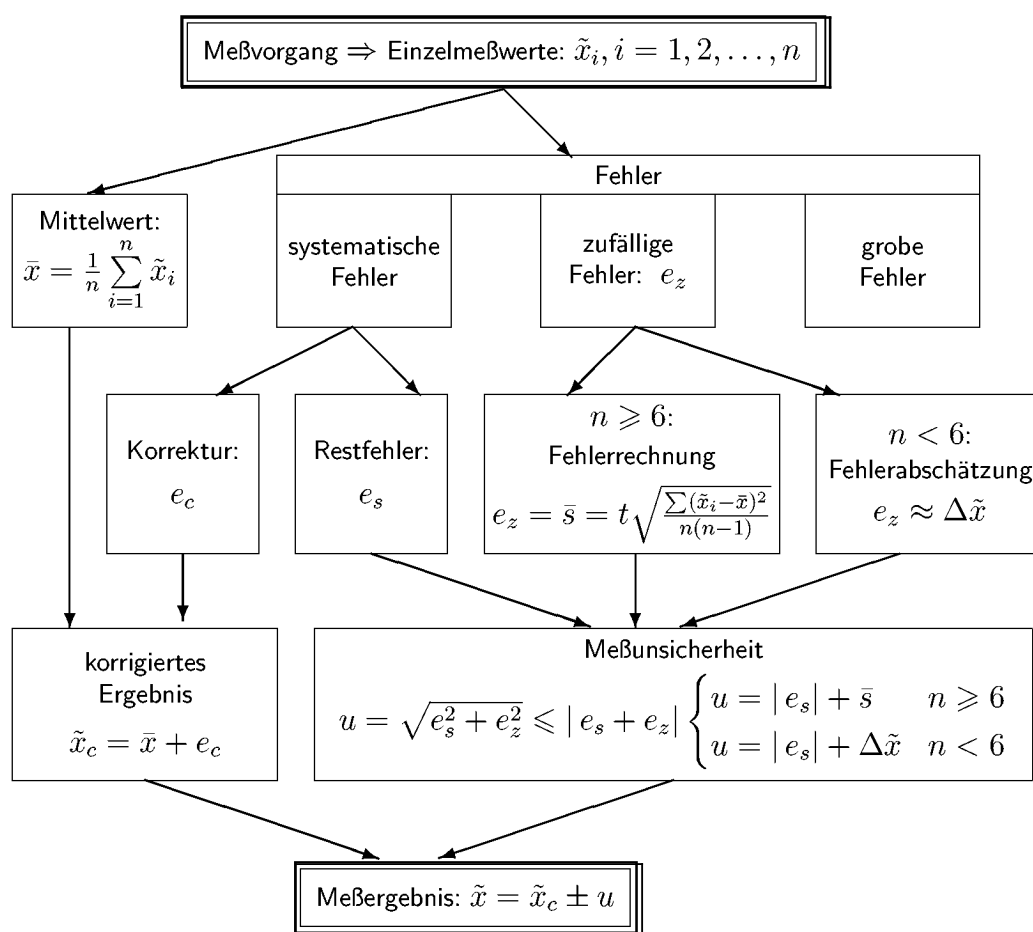


Abb. 4–1 Bestimmung des Meßergebnisses für eine direkt gemessene Größe

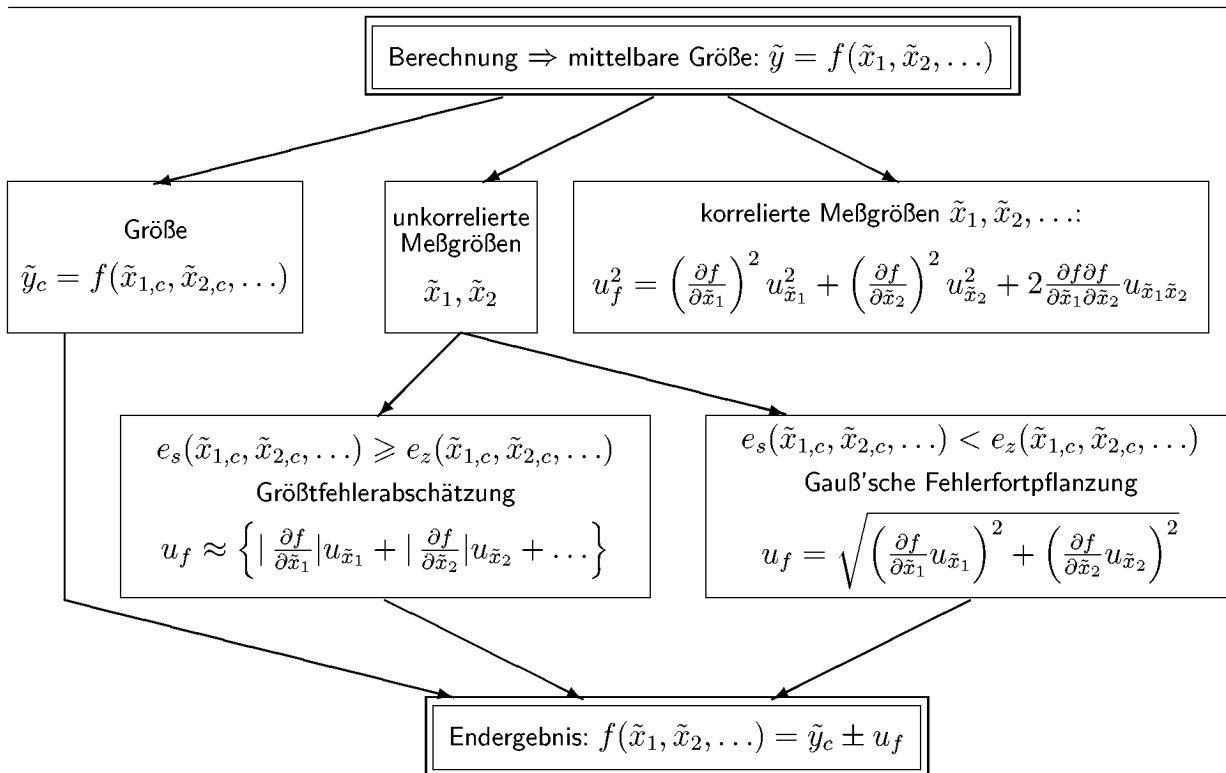


Abb. 4–2 Ermittlung des Meßergebnisses für eine nicht direkt gemessene Größe (Fehlerfortpflanzung) aus den Meßergebnissen für die Einzelmeßgrößen, s. auch Abb. 4–1

Aufgabe einer Fehlerrechnung ist es, aus den streuenden Einzelmeßwerten einer Größe $\tilde{x}_i$ einen wahrscheinlichsten Wert $\tilde{x}_c$ als beste Näherung für den grundsätzlich unbekannten wahren Wert $\tilde{x}_w$ zu bestimmen und eine von den Meßbedingungen abhängige Fehlergröße, die Meßunsicherheit $u_{\tilde{x}}$ anzugeben. Das Resultat der Fehlerrechnung kennzeichnet dann ein Intervall, innerhalb dessen der wahre Wert mit einer kalkulierbaren Wahrscheinlichkeit erwartet werden darf.

In Abbildung 4–1 ist gezeigt, wie das Meßergebnis für eine direkt gemessene Größe unter Berücksichtigung der im Folgenden beschriebenen Fehlerquellen berechnet wird:

Grobe Fehler entstehen durch Irrtümer beim Messen oder Notieren der Werte, Nichtbeachten von größeren äußeren Störeinflüssen, etc. und sind grundsätzlich zu vermeiden.

Systematische Fehler können sehr unterschiedliche Ursachen haben und sind in einer bestimmten Richtung und Größe reproduzierbar wirksam. Es wird zwischen quantitativ erfaßbaren, korrigierbaren Fehlern und Restfehlern, die meist aufgrund der Ungenauigkeit der Meßgeräte entstehen, unterschieden.

Zufällige Fehler entstehen durch meßtechnisch nicht erfaßbare und nicht beeinflussbare Änderungen von Meßgeräten, Meßobjekten und Umwelteinflüssen. Sie genügen statistischen Gesetzen, d.h. es kann mit Hilfe der Statistik der Einfluß der zufälligen Fehler auf das Meßergebnis bestimmt werden, wie z.B. über die Bestimmung des Vertrauensbereiches $\bar{s}$ bei sechs^{4.1} oder mehr Einzelmessungen, in die der Student'sche *t*-Faktor entsprechend der gewählten statistischen Sicherheit einfließt.

Mitunter ergibt sich die gesuchte Größe nicht direkt, sondern aus einem funktionalen Zusammenhang $f = f(\tilde{x}_1, \tilde{x}_2, \dots)$ zwischen verschiedenen direkt meßbaren Größen. Wenn die Mittelwerte und Meßunsicherheiten der Einzelmeßgrößen bekannt sind, erlaubt dann das Fehlerfortpflanzungsgesetz die Berechnung der mittelbaren Größe und ihrer Meßunsicherheit, s. Abb. 4–2. Die Art der Fehlerfortpflanzung

^{4.1} Allgemein als Richtwert festgelegt, da bei einer sehr kleinen Zahl von Einzelmessungen nicht mehr angenommen werden kann, daß diese normalverteilt sind.

wird dabei zunächst davon bestimmt, ob die direkt meßbaren Größen voneinander unabhängig oder ob sie korreliert sind. Bei Vorliegen von unkorrelierten Einzelmeßgrößen ist zu unterscheiden, ob die Anteile der zufälligen Fehler an jeder einzelnen Meßunsicherheit größer als die der systematischen sind. In diesem Fall wird üblicherweise eine Fehlerfortpflanzung nach Gauß und ansonsten eine Berechnung des Größtfehlers durchgeführt.

4.2.2 Fehlerrechnung für die intrazellulären Metabolitkonzentrationen

Die intrazellulären Metabolitkonzentrationen werden über den in Gleichung (3–1) gegebenen Zusammenhang mittelbar aus verschiedenen unkorrelierten Einzelmeßgrößen bestimmt. Bei einigen direkt gemessenen Größen, wie z.B. den Volumina V_N , V_E und V_{Puffer} gelang es nicht, den Anteil der zufälligen Fehler an den Meßunsicherheiten abzuschätzen, bzw. es wurde angenommen, daß der zufällige Anteil kleiner als der des systematischen Fehlers ist ($\Delta\tilde{x} = 0$). Um daher bei der Angabe der Meßunsicherheiten auf der sicheren Seite zu sein, d.h. um einen Maximalfehler abzuschätzen, wurde der Größtfehler aus Gleichung (3–1) wie folgt berechnet:

$$\begin{aligned}
 \Delta\tilde{c} = & \left| F_{\text{verd,LCMS}} \cdot F_{\text{verd,Extr.}} \right| \Delta\tilde{c}_{\text{LCMS}} + \\
 & \left| \tilde{c}_{\text{LCMS}} \cdot F_{\text{verd, Extr.}} \cdot \frac{V_{\text{Puffer}}}{V_{\text{Probe}}^2} \right| \Delta V_{\text{Probe}} + \\
 & \left| \tilde{c}_{\text{LCMS}} \cdot F_{\text{verd,Extr.}} \cdot \frac{1}{V_{\text{Probe}}} \right| \Delta V_{\text{Puffer}} + \\
 & \left| \tilde{c}_{\text{LCMS}} \cdot F_{\text{verd,LCMS}} \cdot \frac{V_E + V_N + V_X}{BTM^2 \cdot V_S \cdot K_E} \right| \Delta BTM + \\
 & \left| \tilde{c}_{\text{LCMS}} \cdot F_{\text{verd,LCMS}} \cdot \frac{V_E + V_N + V_X}{BTM \cdot V_S^2 \cdot K_E} \right| \Delta V_S + \\
 & \left| \tilde{c}_{\text{LCMS}} \cdot F_{\text{verd,LCMS}} \cdot \frac{V_E + V_N + V_X}{BTM \cdot V_S \cdot K_E^2} \right| \Delta K_E + \\
 & \left| \tilde{c}_{\text{LCMS}} \cdot F_{\text{verd,LCMS}} \cdot \frac{1}{BTM \cdot V_S \cdot K_E} \right| \Delta(V_E + \Delta V_N + \Delta V_X).
 \end{aligned} \tag{4-1}$$

Die notwendigen Angaben für die Einzelmeßgrößen waren nur aus den Experimenten MU1 und MU2 bekannt, so daß hier die Größtfehlerabschätzung mit Gleichung (4–1) beispielhaft anhand von MU1 durchgeführt wird, wobei die Werte und die Meßunsicherheiten der direkt gemessenen Größen in Tabelle 4–1 zusammengefaßt sind. Die Meßunsicherheit $\Delta\tilde{c}_{\text{LCMS}}$ setzt sich additiv aus der Unsicherheit bei der Ermittlung des Signalwertes einer Probe, d.h. des Flächenintegrals des LC-MS-Signals $\Delta\tilde{c}_{\text{LCMS,Sig.}}$ und einer Unsicherheit bei der Berechnung der Konzentrationen über die Kalibriergerade $\Delta\tilde{c}_{\text{LCMS,Tol.}}$ zusammen. Da Doppelbestimmungen der Proben durchgeführt wurden, entspricht $\Delta\tilde{c}_{\text{LCMS,Sig.}}$ der Abweichung einer Einzelbestimmung der Konzentration von dem aus den zwei gleichen Proben arithmetisch gebildeten Mittelwert.

^{4.2}Der für *E.coli* K12 bestimmte Wert für das spezifische Zellvolumen K_E wurde aus (Buchholz 2001) übernommen.

Tab. 4–1 Wert und geschätzte Meßunsicherheit der unmittelbaren Größen in Experiment MU1

Größe	Wert ($\tilde{x}$)	Meßunsicherheit ($\Delta\tilde{x}$)	Einheit	Kommentar
c_{LCMS}	–	–	μM	variiert, s. Text
V_{Probe}	100,0	0,8	μl	Volumenfehler der 100 μl Eppendorfpipetten
V_{Puffer}	25,0	0,5	μl	Volumenfehler der 10-50 μl Eppendorfpipetten
BTM	16,28	0,49	g l^{-1}	s. Text
V_S	7,50	0,11	ml	relativer Fehler $\approx 1,4\%$ (Schaefer 1999)
$K_E^{4.1}$	2,78	0,11	g ml^{-1}	relativer Fehler $\approx 4\%$ (Buchholz 2001)
V_E	5,0	0,1	ml	Annahme eines Volumenfehlers von 2%
V_N	$\approx 2,5$	$\approx 0,05$	ml	Annahme eines Volumenfehlers von 2%
V_X	318,0	15,6	μl	relativer Fehler $\approx 4,9\%$ (Buchholz 2001)

Die Meßunsicherheit $\Delta\tilde{c}_{LCMS,Tol.}$ läßt sich aus dem Toleranzbereich^{4.3} der jeweiligen Kalibriergeraden folgendermaßen berechnen (DIN 38 402 Teil 51 1986):

$$\Delta\tilde{c}_{LCMS,Tol.} = \frac{s_{x,y}}{m_{St}} \cdot t \cdot \sqrt{\frac{1}{pr} + \frac{1}{wm} + \frac{(\tilde{c}_{LCMS} - \bar{c}_{Kal})^2}{\sum_{l=1}^{pr} (\tilde{c}_{l,Kal} - \bar{c}_{Kal})^2}}, \quad \bar{c}_{Kal} = \frac{1}{pr} \sum_{l=1}^{pr} \tilde{c}_{l,Kal}. \quad (4-2)$$

m_{St} bezeichnet die Steigung der Kalibriergeraden, $pr=6$ die Zahl der Kalibrierproben $c_{l,Kal}$, $wm=2$ die Zahl der Wiederholungsmessungen der Probe $\tilde{c}_{LCMS}$, und $t_{0,1;4}=2,132$ den Student'schen Faktor unter Voraussetzung eines 90%-igen zweiseitigen Vertrauensbereiches. Die Reststandardabweichung $s_{x,y}$ wird wie folgt berechnet:

$$s_{x,y} = \sqrt{\frac{\sum_{l=1}^{pr} (\tilde{W}_{l,Kal} - W_{l,fit})^2}{pr - 2}},$$

wobei $(\tilde{W}_{l,Kal} - W_{l,fit})$ das Residuum der i -ten Kalibrierprobe, d.h. die Abweichung des Signalwertes $\tilde{W}_{l,Kal}$ von dem über die Kalibriergerade bestimmten Wert $W_{l,fit}$ für die Kalibrierprobe mit der Konzentration $\tilde{c}_{l,Kal}$ ist.

Die Meßunsicherheit der BTM wurde als das arithmetische Mittel der absoluten Residuen einer Ausgleichsgerade für die vorhandenen Wertepaare von OD_{650} und BTM abgeschätzt.

Ergebnisse Die Werte und Meßunsicherheiten für alle intrazellulären Metabolitkonzentrationen, c , aus den Experimenten MU1 und MU2 wurden berechnet. In Abb. 4–3 sind die Ergebnisse der Fehlerrechnung beispielhaft für zwei Metabolite mit einer relativ geringen (Shikimat) bzw. einer relativ großen Meßunsicherheit aufgetragen (G6P+F6P). Insgesamt bewegt sich die zeitlich gemittelte relative Meßunsicherheit:

$$\Delta\bar{c}/\bar{c} = \frac{1}{n} \sum_{i=1}^n \frac{\Delta\tilde{c}_{i,LCMS}}{\tilde{c}_{i,LCMS}}, \quad i = 1, \dots, n : \text{Zeitpunkte},$$

^{4.3}Er beschreibt die Streuung der einzelnen Meßpunkte um die Kalibriergerade und verschwindet nicht, wie der Vertrauensbereich, bei unendlich vielen Messungen.

s. Abb. 4–3 (rechts), meist zwischen 14–34% und ist nur in einigen Ausnahmefällen, d.h. für die Metabolite NAD, ATP und DHQ größer als 60%. Ihre Zusammensetzung aus den einzelnen Fehlerfaktoren, s. Gleichung (4–1), und ihr zeitlicher bzw. konzentrationsabhängiger Verlauf, s. Abb. 4–4, geben Hinweise auf eine verbesserte Durchführung der Experimente. So zeigt sich, daß die Einflüsse der Faktoren V_{Puffer} , V_{Probe} , V_S und $(V_E + V_N + V_X)$ gegenüber den aus der LC-MS- und BTM - bzw. K_E -Bestimmung resultierenden Unsicherheiten eher vernachlässigt werden können. Die Meßunsicherheiten aus der LC-MS-Bestimmung sind insbesondere bei niedrigeren Konzentrationen mit einem Anteil von über 50% am gesamten Größtfehler dominant. Hier hätte eine Verschiebung des Arbeitsbereiches der Kalibrier-

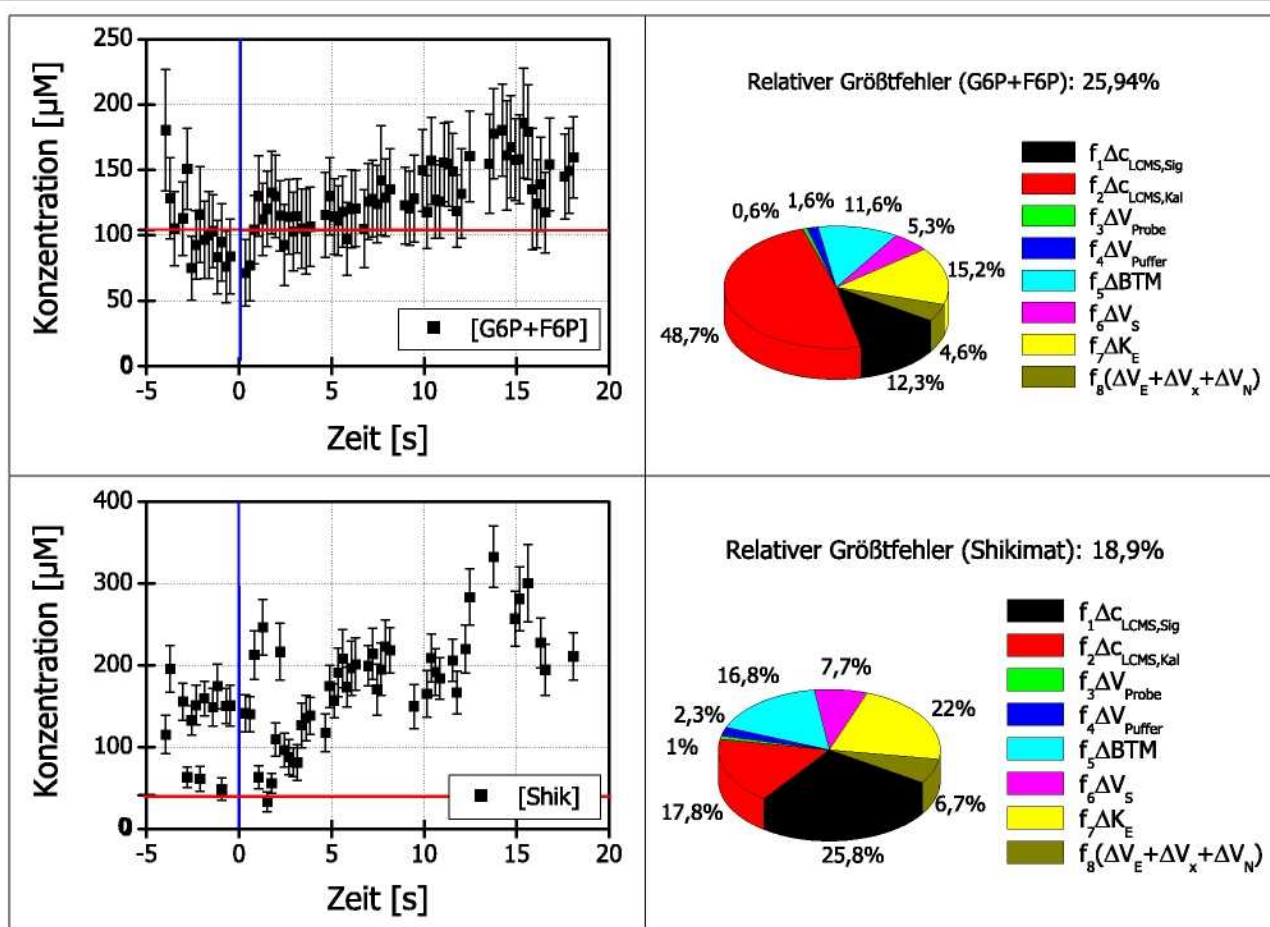


Abb. 4–3 Meßergebnisse (links) und Zusammensetzung des zeitlichen gemittelten Größtfehlers (rechts) für zwei ausgewählte Metabolite aus Experiment MU1 (f: Fehleranteile); die horizontalen Linien zeigen die mit den Verdünnungsfaktoren verrechneten Bestimmungsgrenzen, s. Abschnitt 4.3, die senkrechten Linien den Zeitpunkt der Pulsaufgabe an

geraden, s. dazu Abschnitt 4.3, und eine Dreifach- statt der vorgenommenen Doppelbestimmung der Meßsignale die Meßunsicherheit der LC-MS-Bestimmung bereits deutlich erniedrigen können.

Ebenso läßt sich aus den mit ansteigenden Konzentrationen stark zunehmenden Anteilen der Unsicherheiten der BTM - bzw. K_E -Bestimmung ableiten, daß diese Größen für jedes Experiment neu und besonders sorgfältig bestimmt werden sollten.

Generell resultiert aus einer Größtfehlerabschätzung immer nur der maximale Wert für das Intervall, innerhalb dessen sich der wahre Meßwert mit einer hohen Wahrscheinlichkeit befindet, d.h. das Intervall kann mit steigendem Informationsgehalt aus experimentellen Messungen stark verringert werden. Bei

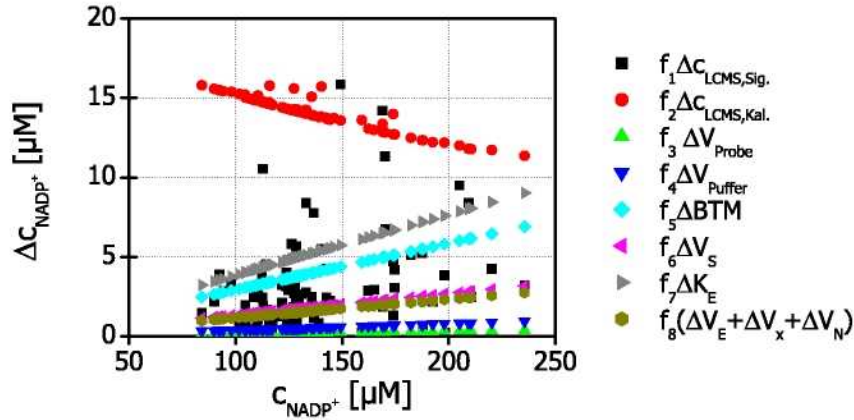


Abb. 4–4 Konzentrationsabhängigkeit des Größtfehlers für den Metaboliten NADP aus Experiment MU1

einer sechsfachen Bestimmung aller Einzelmeßgrößen hätte die Anwendung der Gauß'schen Fehlerfortpflanzung, bzw. bei einer mehrmaligen Wiederholung desselben Experimentes die Berechnung des Vertrauensbereiches der intrazellulären Metabolitkonzentrationen wesentlich geringere Meßunsicherheit ergeben können. In (Schmitz 2000) wurde z.B. bei einem unter nicht vergleichbaren Bedingungen durchgeführten Experiment für die Quantifizierung des Metaboliten NAD ($\tilde{c}_{NAD} \approx 12 \text{ mM}$) eine relative Standardabweichung von 6,14% anhand von sechs vermessenen Proben ermittelt.

Daher wurden die Resultate der Fehlerrechnung außer zu der bereits erfolgten Analyse der einzelnen Fehlerquellen weiterhin nur dazu verwendet, um einzelne Metabolite mit einer sehr hohen Meßunsicherheit von der Modellierung auszuschließen, die Ergebnisse der Datenaufbereitung, s. Abschnitt 4.5, zu beurteilen und beispielhaft weitere Kennzahlen für die Bewertung der Anpassungsgüte der Modelle zu berechnen.

4.3 Nachweis-, Erfassungs- und Bestimmungsgrenzen

Standardisierte Kriterien zur Bewertung traditioneller analytischer Arbeitsmethoden liegen in den Nachweis-, Erfassungs- und Bestimmungsgrenzen vor, s. (DIN 32 645 1994).

Die Nachweisgrenze (NG) ist die Entscheidungsgrenze für das Vorhandensein eines Stoffes und bezeichnet die geringste Probenkonzentration, die mit einer gewissen Irrtumswahrscheinlichkeit α qualitativ von einer Blindprobe unterschieden werden kann. Sie berechnet sich folgendermaßen aus den vermessenen Kalibrierproben und der Kalibriergeraden:

$$\tilde{c}_{NG} = \frac{s_{x,y}}{m_{St}} \cdot t \cdot \sqrt{\frac{1}{pr} + \frac{1}{wm} + \frac{\bar{c}_{Kal}^2}{\sum_{l=1}^{pr} (\tilde{c}_{l,Kal} - \bar{c}_{Kal})^2}}, \quad (4-3)$$

wobei wm die Zahl der Wiederholungsmessungen der Kalibrierproben bezeichnet und alle anderen Größen die gleiche Bedeutung wie in Gleichung (4–2) besitzen. Der Faktor der Student'schen Verteilung wird entsprechend eines 99%-igen einseitigen Vertrauensbereiches gewählt. Die Erfassungsgrenze (EG) ist der kleinste Gehalt einer gegebenen Probe, bei dem mit der Wahrscheinlichkeit $1 - \beta$ ein Nachweis möglich ist. Es gilt: $\tilde{c}_{EG} = 2 \cdot \tilde{c}_{NG}$, wenn, wie in diesem Fall, $\alpha = \beta$ gesetzt wird, s. (DIN 32 645 1994).

An der Bestimmungsgrenze (BG) kann der Gehalt einer Probe mit einer vorgegebenen Wahrscheinlichkeit quantitativ bestimmt werden. Ihr Wert wird iterativ berechnet:

$$\Delta \tilde{c}_{BG} = \frac{s_{x,y}}{m_{St}} \cdot t \cdot \sqrt{\frac{1}{pr} + \frac{1}{wm} + \frac{(\tilde{c}_{BG} - \bar{c}_{\text{Kal}})^2}{\sum_{l=1}^{pr} (\tilde{c}_{l,\text{Kal}} - \bar{c}_{\text{Kal}})^2}}, \quad \tilde{c}_{BG} = k_u \cdot \Delta \tilde{c}_{BG}, \quad (4-4)$$

indem $\tilde{c}_{BG,0} = t \cdot \tilde{c}_{NG}$ als Startwert, eine Ergebnisunsicherheit von 33,3%, d.h. $k_u=3$, und ein Student'scher t -Faktor für einen 99%-igen zweiseitigen Vertrauensbereich angenommen wird.

In Tab. 4–2 sind die Ergebnisse für die NG, EG und BG aufgelistet, wobei die EG und NG jeweils mit den in (DIN 32 645 1994) angegebenen Faktoren 0,6 bzw. 2,87 für sechs Proben und ein Signifikanzniveau von $\alpha=5\%$ multipliziert wurden, um ihre Vertrauensbereiche zu ermitteln. Nach Berücksichtigung eines Verdünnungsfaktors von $F = F_{\text{verd,LCMS}} \cdot F_{\text{verd,Extr.}} = 29,881$ ergeben sich aus den Werten für die BG die in Abb. 4–3 eingezeichneten horizontalen Linien, unterhalb derer die intrazellulären Konzentrationen nicht mehr quantitativ angegeben werden sollten. Wie in dem Fall der Summenkonzentration von

Tab. 4–2 Bestimmungs-, Nachweis- und Erfassungsgrenzen für die Kalibrierfunktionen aus Experiment MU1

Substanz	BG [μ M]	NG [μ M]			EG [μ M]		
G6P+F6P	3,477	$0,727 \leq$	$1,211 \leq$	3,476	$1,453 \leq$	$2,422 \leq$	6,951
FBP	2,898	$0,571 \leq$	$0,951 \leq$	2,730	$1,141 \leq$	$1,902 \leq$	5,459
DHAP+GAP	2,508	$0,490 \leq$	$0,816 \leq$	2,342	$0,979 \leq$	$1,632 \leq$	4,684
2PG+3PG	2,375	$0,459 \leq$	$0,766 \leq$	2,197	$0,919 \leq$	$1,531 \leq$	4,394
PEP	2,860	$0,572 \leq$	$0,953 \leq$	2,736	$1,144 \leq$	$1,906 \leq$	5,470
6PG	0,453	$0,076 \leq$	$0,127 \leq$	0,364	$0,152 \leq$	$0,254 \leq$	0,729
R5P	4,765	$1,075 \leq$	$1,792 \leq$	5,142	$2,150 \leq$	$3,583 \leq$	10,283
DAH	141,111	$39,155 \leq$	$65,258 \leq$	187,290	$78,310 \leq$	$130,516 \leq$	374,581
3DHQ	5,812	$1,296 \leq$	$2,160 \leq$	6,198	$2,591 \leq$	$4,319 \leq$	12,396
DHS	2,776	$0,551 \leq$	$0,919 \leq$	2,638	$1,103 \leq$	$1,838 \leq$	5,275
Shik	1,376	$0,247 \leq$	$0,412 \leq$	1,182	$0,494 \leq$	$0,823 \leq$	2,362
S3P	4,124	$0,900 \leq$	$1,500 \leq$	4,304	$1,799 \leq$	$2,999 \leq$	8,607
NADP	3,329	$0,686 \leq$	$1,143 \leq$	3,282	$1,372 \leq$	$2,287 \leq$	6,564
NAD	6,415	$1,635 \leq$	$2,726 \leq$	7,822	$3,271 \leq$	$5,451 \leq$	15,644
ATP	7,775	$2,130 \leq$	$3,550 \leq$	10,189	$4,260 \leq$	$7,100 \leq$	20,377
ADP	7,155	$1,910 \leq$	$3,184 \leq$	9,137	$3,382 \leq$	$6,367 \leq$	18,273
AMP	76,579	$12,696 \leq$	$22,273 \leq$	83,080	$26,728 \leq$	$44,547 \leq$	127,850

G6P und F6P liegen die Konzentrationen fast aller Metabolite in der Nähe der BG, was weniger auf die Empfindlichkeit des Meßgerätes, die z.B. über das Signal-Rausch-Verhältnis ausgedrückt werden kann (Oldiges 2004), als auf die besondere Arbeitstechnik bei der Erstellung der Kalibriergeraden zurückzuführen ist: da die Effekte der Probenmatrix auf die LC-MS-Bestimmung^{4,4} berücksichtigt werden sollten, wurden auf Proben aus dem Fließgleichgewichtszustand vor dem Puls unterschiedliche Volumina einer Maßlösung der zu bestimmenden Metabolite gegeben, um die Kalibrierproben zu erzeugen (Oldiges 2004). Der niedrigste Konzentrationswert der Kalibriergeraden, der hier meist in etwa der BG gleichkommt, entspricht somit gerade der Konzentration des Analyten im Fließgleichgewichtszustand, was zu Schwierigkeiten bei der Quantifizierung von kleineren Konzentrationswerten führt, welche möglicherweise nach dem Substrat-Puls auftreten. Eine Verbesserung der Ergebnisse könnte z.B. über die Quantifizierung der Metabolite in sämtlichen Proben mit Hilfe der Methode der Standardaddition erzielt werden, was jedoch bei einer Probenanzahl von ca. 160 recht zeitaufwendig wäre.

^{4,4}Unter der Probenmatrix werden alle neben dem Analyten zusätzlich gelösten und evtl. störenden Begleitkomponenten verstanden, die das Meßergebnis oft erheblich beeinflussen können, wie z.B. hier vorwiegend die aufgrund der Neutralisation enthaltene hohe Salzkonzentration.

4.4 Standardisierung der Datenerfassung – Software PulseSTAT

Es erweist sich stets als sinnvoll für die Modellierung, wenn für alle Experimente einheitliche Kriterien definiert werden, anhand derer die Meßdaten bewertet oder verschiedene Experimente miteinander verglichen werden können, wie z.B. hier die Angabe der Meßunsicherheiten und der BG.

Die Meßdaten und die Kriterien sollten dann möglichst schnell nach der Durchführung des Experimentes fehlerfrei vorliegen, um den Beginn der Modellierung nicht unnötig zu verzögern. Aus dem gleichen Grund ist eine übersichtliche und vollständige Angabe aller für die Modellierung benötigten Meßparameter, wie z.B. der beim Substrat-Puls injizierten Glucose-Menge, notwendig.

Dafür wurde auf der Basis der gesamten Rohdaten aus den Experimenten MU1 und MU2 im Rahmen dieser Arbeit die Software PulseSTAT entwickelt. PulseSTAT ist eine VBA-Anwendung (VBA: Visual Basic for Application), die in Microsoft Excel (Versionen ab Excel 2000) lauffähig ist. Alle im Verlauf der Puls-Experimente erfaßten Daten können ausgehend vom Startfenster, s. Abb. 4–5, direkt in entsprechend vorbereitete Excel-Tabellenblätter eingetragen und danach automatisch ausgewertet werden.

Größe	Wert	SI-Einheit	Fehler
OD (SI-Einheit entspricht der BTM in g/l)	50,2	16,28	0,497940705
Probenvolumen (V_E) [ml]	7,5	0,0075	0,000105
Spezifisches Zellvolumen (K_E) [μ l/mg]	2,78	0,00278	0,0001112
Interzelluläres Volumen (V_X) [μ l]	318	0,000318	0,000015582
Volumen Extraktionslösung (V_E) [ml]	5	0,005	0,00003
Probenvolumen für die LC-MS-Bestimmung (V_{Probe}) [μ l]	100	0,0001	0,0000008
Volumen der Pufferlösung für die LC-MS-Best. (V_{Puffer}) [μ l]	25	0,000025	0,0000005
Probenzahl	83	%	%

Buttons: Biotrockenmasse, Verdünnung, LCMS, Gesamtfehler, Datentest

Abb. 4–5 Startfenster des Auswertungswerkzeuges PulseSTAT; allgemeine Parameter des Experimentes werden festgelegt, verschiedene Buttons führen zu weiteren Tabellenblättern zur Auswertung aller Abschnitte eines Puls-Experimentes

Im Folgenden werden die Optionen zur Auswertung erläutert, die über den Aufruf der verschiedenen Tabellenblätter im Startfenster, s. Abb. 4–5, möglich sind:

1. Biotrockenmasse:

Eingabe und lineare Regression der OD_{650} und BTM -Wertepaare; Eintragen des Verlaufs der OD_{650} über der Fermentationszeit; Berechnung des Verlaufes der BTM über die Fermentationszeit entsprechend der OD_{650} - BTM -Relation; Ermittlung der exponentiellen Wachstumsphase mit der entsprechenden Wachstumsrate und ihrer Meßunsicherheit; Berechnung der BTM zum Zeitpunkt des Glucose-Pulses dem angegebenen OD_{650} -Wert sowie mehrere Optionen zur Bestimmung der Meßunsicherheit für diesen BTM -Wert;

2. *Verdünnung:*

Eingabe der jeweils benötigten Neutralisationsvolumina, Angabe eines Faktors für deren relative Meßunsicherheit; Berechnung der Verdünnungsfaktoren und der Meßunsicherheiten für die Größen V_{Probe} , V_{Puffer} , BTM , V_S , K_E , $(V_X + V_E + V_N)$ unter Berücksichtigung der Angaben im Startfenster, s. Abschnitt 4.2;

3. *LC-MS:*

Eingabe der Volumina und Konzentration der zugegebenen Standardadditionslösung, bzw. der Konzentrationswerte bei einer externen Kalibrierung mit zugehörigen Signalflächen aus den Mehrfachbestimmungen der Kalibrierproben; lineare oder quadratische Regression, Ermittlung der Toleranzbereiches der Kalibrierfunktion, Berechnung der BG, NG und EG aus den Kalibriergeraden; Eintragen der Signalwerte aus Mehrfachbestimmungen der Analysenlösung; Berechnung der Konzentrationen und der Meßunsicherheiten $\Delta\tilde{c}_{\text{LCMS,Kal}}$ und $\Delta\tilde{c}_{\text{LCMS,Sig}}$, wie in Abschnitt 4.2 beschrieben;

4. *Gesamtfehler:*

Eintragen der Zeitachse, Berechnung der intrazellulären Metabolitkonzentrationen und ihrer Meßunsicherheiten aus den Angaben der anderen Tabellenblätter, s. Abschnitt 4.2; Export der Zeitachse, der Konzentration $\tilde{c}$ und deren Meßunsicherheit $\Delta\tilde{c}$ in ein gesondertes Tabellenblatt, in dem diese Angaben für beliebig viele Metabolite gesammelt werden können; automatische Erzeugung der in Abb. 4–3 und Abb. 4–4 gezeigten Diagramme;

5. *Datentest:*

bei Eingabe von Wertepaaren: Durchführung einer linearen bzw. quadratischen Regression; Prüfen der Annahme der Linearität gemäß dem Mandel-Test (Graf et al. 1966); F-Test zur Erkennung der Ausreißerwerte in der Kalibrierfunktion (Graf et al. 1966); graphische Auftragung sowie R/s-Test der Residuen auf Normalverteilung für die lineare bzw. quadratische Ausgleichsfunktion (Sachs 2004, S. 424);

bei Eingabe einer Datenreihe: Berechnung des Mittelwertes und der Standardabweichung, Durchführung eines Trendtests nach Neumann (Sachs 2004, S. 481), eines Ausreißertests nach Dixon (Sachs 2004, S. 362)^{4.5};

Die Ergebnisse, d.h. die Größen und Meßunsicherheiten, sowie die BG, NG und EG werden automatisch aus den eingetragenen Rohdaten berechnet, wodurch sich die zur Datenauswertung benötigte Zeit wesentlich verkürzt^{4.6}, die Möglichkeit für Rechenfehler minimiert wird, und die Rohdaten parallel zur Durchführung der Experimente ausgewertet werden können, was eine Rückkopplung auf die Experimente erlaubt. Die vollständige Protokollierung der Experimente wird aufgrund der vorgefertigten Tabellenblätter für die Rohdaten erheblich erleichtert. Darüber hinaus können die Tabellenblätter exportiert werden, so daß die notwendigen Daten direkt ohne weitere Rechenschritte ausgelesen und im Zuge der Modellierung weiterverarbeitet werden können.

4.5 Glättung der Daten

Da bei der Parameteranpassung eines Modells die Konvergenz eines Optimierungsalgorithmus wesentlich von der Datenqualität beeinflusst wird, s. Abschnitt 3.4.1, wurde entschieden, der Modellierung in

^{4.5}Die Werte der Meßreihe werden ihrer Größe nach aufsteigend ($x_1 < x_2 < \dots < x_k$) bzw. abfallend ($x_1 > x_2 > \dots > x_k$) geordnet. Bei dem vorliegenden Stichprobenumfang von ca. 16 wird dann die Prüfgröße: $|(x_1 - x_3)/(x_1 - x_{k-2})|$ berechnet. Überschreitet die Prüfgröße den kritischen Wert für ein gewähltes Signifikanzniveau α , s. (Sachs 2004, Tab. 97), so ist der kleinste bzw. der größte Wert x_1 ein Ausreißer der Meßreihe. Für diese Arbeit ist α auf 0,1 gesetzt worden.

^{4.6}Die komplette Auswertung eines Puls-Experimentes läßt sich nun in ein bis zwei Tagen im Gegensatz zu den bisher benötigten drei Wochen durchführen.

geeigneter Form aufbereitete Daten zugrunde zu legen. Dazu wurden zunächst Ausreißerwerte eliminiert, d.h. Datenpunkte mit einer relativ hohen Meßunsicherheit oder mit einer gemäß dem Dixon-Test ziemlich großen Abweichung von der Normalverteilungsfunktion um den Fließgleichgewichtswert. Danach wurden verschiedene Methoden zur Glättung der Daten getestet, die im Folgenden vorgestellt werden.

4.5.1 FFT-Filterung

In vorbereitenden Schritten wurde eine Interpolation der Datenreihen vorgenommen, um Werte zu äquidistanten Zeitpunkten zu erhalten, und anschließend die Zeitdomäne des Spektrums über eine Fast-Fourier-Transformation (FFT) in eine Frequenzdomäne überführt s. (Bronstein und Semendajew 1989, S. 616 ff). Zur Glättung der Daten wurde ein spezieller Tiefpaßfilter verwendet. Hierbei handelt es sich um eine Parabel, die für eine Frequenz von Null ein Maximum von 1 einnimmt und bei einer Grenzfrequenz oberhalb derer alle Frequenzen herausgefiltert werden sollen, auf Null abfällt. Der Krümmungsparameter der parabolischen Sperrfunktion wird durch die Gesamtzahl der Meßpunkte und eine einstellbare Anzahl der jeweils betrachteten Punkte bestimmt, wobei der Grad der Glättung um so größer ist, je mehr Punkte zeitgleich betrachtet werden. Nach der Filterung werden die geglätteten Trajektorien über eine Rücktransformation der Frequenz- in die Zeitdomäne erhalten. FFT-Filterungen

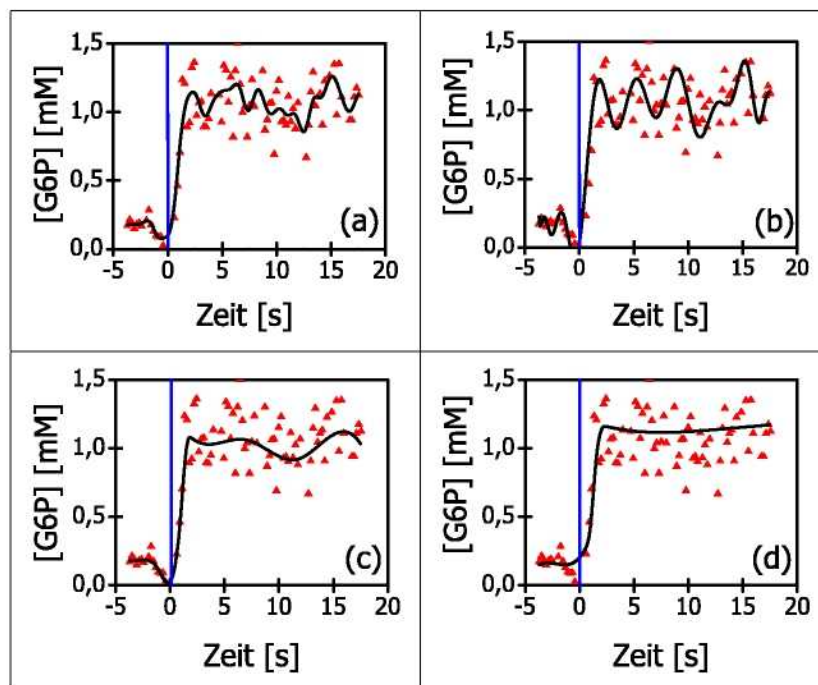


Abb. 4–6 Glättung der G6P-Trajektorie aus Experiment WT1: (a) FFT-Filterung, (b) Approximation mit Tschebyscheff-Polynomen, (c) Approximation mit gewichteten kubischen B-Splines und Wahl des Glättungsfaktors bzw. (d) der inneren Stützstellen

können nach Implementierung entsprechender Routinen, wie z.B. aus der NAG C Bibliothek^{4,7}, rasch und automatisiert ausgeführt werden oder sind bereits in gängigen Kalkulationsprogrammen, wie z.B. Origin 6.1 (OriginLab Corporation, 1991-2000) implementiert. Der jeweils vorzugebende Schwellwert des Tiefpaßfilters läßt sich jedoch nicht aus biologischen bzw. experimentellen Gegebenheiten ableiten, was evtl. zu Verfälschungen der Glättungsergebnisse führen kann, s. Abb. 4–6 (a). Daher sollten die gefilterten Daten mit den originalen Meßdaten, z.B. auf zusätzlich auftretende Schwingungen, verglichen werden.

^{4,7}<http://www.nag.com/numeric/CL/CLdescription.asp>

4.5.2 Polynomische und Splineapproximationen

Neben der im letzten Abschnitt dargestellten Filterung der Daten wurden Methoden zur Approximation der Trajektorien mit unterschiedlichen Näherungsfunktionen verwendet.

Als einfachste Variante wurde zunächst eine polynomische Interpolationsfunktion zur Approximation der Datenpunkte getestet. Hierfür wurde aus einer Zahl von weiteren, in Frage kommenden Polynomfunktionen eine Reihenentwicklung von Tschebyscheff-Polynomen ausgewählt, da diese aufgrund ihrer extremalen Eigenschaften als besonders geeignet für numerische Approximationsprobleme beschrieben werden, s. z.B. (Hanke Bourgeois 2002, S. 284 ff)^{4,8}. Das Interpolationsverfahren wurde in eine Maple-Routine (Maple 8.01 Waterloo Maple Inc. 1981-2002) implementiert, die die Maple-eigene Funktion „chebyshev“ aufruft, so daß die notwendige Transformation der unabhängigen Variablen des zu untersuchenden Zeitintervalls auf den Definitionsbereich der Polynome ($\mathbb{D} = [-1, 1]$) und die darauffolgende Polynomanpassung mit einer einstellbaren Genauigkeit automatisch durchgeführt werden kann. Bei dem Test des Verfahrens mit den vorliegenden Meßdaten wurde jedoch häufig festgestellt, daß die Polynome bei Wahl einer geringen Genauigkeit die Daten nicht hinreichend genau wiedergeben konnten und bei Verwendung einer höheren Genauigkeit zu oszillieren begannen, wobei die Extrema meist nur in geringer qualitativer Übereinstimmung mit denen der experimentellen Trajektorien standen, s. Abb. 4–6 (b).

In dieser Hinsicht geeigneter erschienen daher Approximationen mit Funktionen aus stückweise in Stützstellen zusammengesetzten Polynomen niedrigem Grades, den Splinefunktionen (Hanke Bourgeois 2002, S. 355). Dazu wurden zwei verschiedene Verfahren auf der Basis von Splinefunktionen, die aus gewichteten, kubischen B-Splines bestehen (Hartung und Elpelt 1995), verwendet, s. (Froemel 2003) für die zugrundeliegende Theorie und die Implementierung der entsprechenden NAG C-Routinen^{4,7} e02bac bzw. e02bec. Bei beiden Verfahren wurden die Wichtungskoeffizienten der kubischen B-Splines so bestimmt, daß die L_2 -Norm (euklidische Norm) als Maß für den gewichteten Abstand der Splinefunktion von allen Datenpunkten minimiert wurde.

Bei dem zuerst implementierten Verfahren (Routine: „b-spline-glatt“) konnte ein Glättungsfaktor als Abbruchskriterium der Optimierung angegeben werden, wodurch eine möglichst glatte Splinefunktion unter der Nebenbedingung, daß die minimierte L_2 -Norm kleiner oder gleich dem Glättungsfaktor sein sollte, bestimmt wurde, s. Abb. 4–6 (c). Bei Verwendung dieses Verfahrens ist eine deutliche Verbesserung der Resultate bezogen auf die der polynomischen Approximation ersichtlich. Es zeigte sich jedoch bei Betrachtung aller Metabolittrajektorien, daß rasche Konzentrationsänderungen, wie sie meist unmittelbar auf den Substrat-Puls folgen, nicht ganz exakt wiedergegeben werden können.

Daher wurde ein weiteres Verfahren implementiert, das erlaubte, die Zahl und die Lage der Stützstellen zu wählen, und die L_2 -Norm in Abhängigkeit von den gewählten Stützstellen zu minimieren (Routine: „b-spline-stuetz“). Dadurch konnten an Stellen, an denen die Daten ein sehr dynamisches Verhalten vermuten ließen, mehr und in Bereichen schwächerer Variation der Datenreihen weniger Stützstellen gewählt werden. Dies führte im Allgemeinen zu einer leicht verbesserten Wiedergabequalität, s. Abb. 4–6 (d), die höchstens durch unerwünschte Schwingungen der Splinefunktion bei der Wahl von zu vielen Stützpunkten beeinträchtigt wurde.

Bei einem abschließenden Vergleich aller getesteten Glättungsfunktionen läßt sich auf Basis der gewonnen Erkenntnisse bemerken, daß die FFT-Filterung dann verwendet werden sollte, wenn die Streuungen der Meßdatenpunkte relativ groß und damit die zugrundeliegenden deterministische Verläufe nicht

^{4,8}Die $n + 1$ Extrema der n Tschebyscheff-Polynome konzentrieren sich an den Enden des Intervalls, wodurch der Interpolationsfehler im Gegensatz zu Verfahren auf der Basis von interpolierenden Polynomen mit äquidistanten Stützstellen deutlich verringert wird.

unbedingt eindeutig erkennbar sind, während ansonsten eine Splineapproximation über kubische B-Splines mit Wahl der Stützstellen vorzuziehen ist. Daher wurden die Meßdaten aus den Experimenten MU1 und MU2 FFT-gefiltert und die Daten aus Experiment WT1 mit der Routine „b-spline-stuetz“ approximiert, bevor die Modelle an die entsprechend aufbereiteten Daten angepaßt wurden.

4.6 Plausibilitätstests

Die bisher durch die Fehlerrechnung, Ausreißereliminierung und Glättung der Trajektorien durchgeführte Vorbehandlung der Meßdaten läßt noch keine Rückschlüsse auf etwaige Inkonsistenzen in den vorliegenden Datensätzen zu. Um diesbezüglich den gesamten Datensatz zu prüfen bzw. sich offensichtlich im Widerspruch zu den übrigen Trajektorien befindliche Konzentrationsverläufe zu identifizieren, wurden Plausibilitätstests in Kombination mit den in Abschnitt 4.7 beschriebenen Methoden herangezogen.

Dazu wurden zunächst die gemessenen intrazellulären Metabolitkonzentrationen aus dem Fließgleichgewichtszustand mit veröffentlichten Literaturwerten aus ähnlichen Experimenten verglichen, s. Tab. 4–3 für eine Zusammenfassung. Bei Betrachtung der unteren zwei durch gestrichelte Linien abgetrennten Teile der Tabelle zeigt sich, daß größere Konzentrationsunterschiede weniger auf die Verwendung verschiedener Mikroorganismen als auf die Versuchsdurchführung zurückzuführen sind. Denn die Ergebnisse für die mit ähnlichen Arbeitsmethoden untersuchten *E.coli*- und *S.cerevisiae*-Organismen im mittleren Teil der Tabelle sind im Rahmen von Größenordnungen vergleichbar, während für die genetisch mit *S.cerevisiae* CBS 7336 fast identischen Hefezellen im unteren Teil der Tabelle deutlich geringere Konzentrationswerte ermittelt wurden, die jedoch untereinander wiederum in guter Übereinstimmung stehen. Die nähere Untersuchung der betreffenden Arbeitsmethoden ergibt, daß offenkundig sowohl die Wahl des Extraktionsverfahrens als auch die der Quantifizierungsmethode die Ergebnisse beeinflusst. In Teil 2 der Tabelle wurden nämlich deutlich abweichende Werte für G6P in *E.coli* K12 W3110 und PEP in *S.cerevisiae* CBS 7336 auf der Basis von zwei unterschiedlichen Extraktionsverfahren bei ansonsten gleicher Versuchsdurchführung bestimmt. Dies könnte evtl. mit einer mangelnden Säurestabilität der Metabolite beim Perchlorsäure-Aufschluß oder mit der mangelnden Unterscheidbarkeit zwischen intra- und extrazellulären Metabolitkonzentrationen beim N₂-Aufschluß begründet werden. Aus Teil 1 der Tabelle, bzw. Anhang C, geht hervor, daß die Konzentrationen aus den Experimenten MU1 und MU2, bei denen die Metabolite zwar auf dieselbe Weise extrahiert, aber über eine LC-MS-Methode anstatt über enzymatische Assays quantifiziert wurden, meist um einen Faktor 10 niedriger sind als im Versuch WT1. Eine Begründung hierfür könnte sein, daß mit gekoppelten Enzymassays aufgrund von im Assay vorhandenen Effektoren häufig überhöhte Konzentrationswerte bestimmt werden, s. (Kruger 1995; Oldiges und Takors 2004) für eine ausführliche Diskussion.

Insgesamt kann allerdings allein aus einer Zusammenstellung wie in Tabelle 4–3 nicht generell abgeleitet werden, daß die Meßergebnisse von einigen Metaboliten verworfen werden sollten, oder bestimmte Extraktions- bzw. Quantifizierungsverfahren weniger geeignet für die Durchführung von Substrat-Puls-Experimenten sind. Dazu wären vertiefte experimentelle Untersuchungen der einzelnen Arbeitsschritte unter gleichen Bedingungen notwendig gewesen, die bisher kaum in der Literatur gefunden werden können^{4,9}.

In Analogie zu den Konzentrationswerten im Fließgleichgewichtszustand wurden auch die qualitativen Verläufe der Konzentrationen nach dem Substrat-Puls mit denen aus veröffentlichten Experimenten verglichen und auf Plausibilität geprüft. So unterscheidet sich der Konzentrationsverlauf verschiedener Metabolite auf den Glucose-Puls hin in den drei durchgeführten Experimenten stark, wobei sich insgesamt nur im Experiment WT1 nach dem Puls deutliche Veränderungen in den Konzentrationswerten

^{4,9}Dabei sollten insbesondere die Fermentationsbedingungen gründlich untersucht werden, da die Meßergebnisse offenkundig auch in höherem Ausmaß von der gewählten Fermentationsstrategie abhängig sind, s. (Oldiges 2004).

zeigen. Z.B. wird das im Allgemeinen gefundene, charakteristische Sättigungsverhalten des Metaboliten G6P sowie der typische Abfall der PEP-Konzentration unmittelbar nach dem Puls (Visser et al. 2000; Chassagnole et al. 2002) bei den Experimenten MU1 und MU2 nicht beobachtet.

Tab. 4–3 Vergleich dieser Arbeit zugrundeliegender, repräsentativer, intrazellulärer Metabolitkonzentrationen mit Literaturwerten für den Glucose-limitierten Fließgleichgewichtszustand vor dem Substrat-Puls

Organismus	G6P	F6P	FBP	3PG	2PG	PEP	ATP
<i>E.coli</i> K12 DSM498 ^a	0,19 mM	0,16 mM	2,68 mM	2,33 mM	2,36 mM	0,92 mM	0,44 mM
<i>E.coli</i> LJ110 4pF20 ^b	0,10 mM	— ^c	0,05 mM	0,12 mM	— ^c	0,12 mM	0,66 mM
<i>E.coli</i> LJ110 4pF78 ^d	0,11 mM	— ^c	0,04 mM	0,15 mM	— ^c	0,15 mM	0,04 mM
<i>E.coli</i> K12 W3110 ^e	3,48 mM	0,60 mM	0,27 mM	—	—	2,67 mM	4,27 mM
<i>S.cerevisiae</i> CBS 7336 ^f	0,93 mM	0,17 mM	0,11 mM	1,33 mM	—	0,05 mM	2,1 mM ^g
<i>S.cerevisiae</i> CBS 7336 ^h	1,25 mM	—	—	—	—	—	3 mM ^g
<i>E.coli</i> K12 W3110 ⁱ	—	—	—	—	—	2,75 mM	—
<i>S.cerevisiae</i> CEN.PK113-7D ^j	24,76 μ M	3,53 μ M	1,14 μ M	25,00 μ M	— ^f	12,67 μ M	—
<i>S.cerevisiae</i> CEN.PK113-7D ^{k,l}	~50 μ M	~11 μ M	1-2 μ M	~37 μ M	— ^f	~31 μ M	250 μ M
<i>S.cerevisiae</i> CEN.PK113-7D ^{m,l}	25-73 μ M	8-20 μ M	3-11 μ M	34-73 μ M	— ^f	22-50 μ M	176 μ M

^aWT1 (Buchholz 2001): $D=0,125 \text{ h}^{-1}$, $BTM \approx 10 \text{ g l}^{-1}$, Quenching: -40°C MeOH (60% v/v), Extraktion: -18°C HClO_4 (43,8% w/v) mit anschließender Neutralisation (gesättigte K_2CO_3 -Lösung), Analytische Methode: enzymatische Assays (G6P, F6P, FBP, 3PG, 2PG, PEP), ESI-LC-MS (ATP).

^bMU1 (Oldiges 2004): $D \approx 0 \text{ h}^{-1}$ (Fed-Batch), $BTM = 16, 28 \text{ g l}^{-1}$, Quenching: -40°C MeOH (60% v/v), Extraktion: -18°C HClO_4 (43,8% w/v) mit anschließender Neutralisation (gesättigte K_2CO_3 -Lösung), Analytische Methode: LC-MS-MS.

^cDie Summenkonzentration der Metabolite wurde quantifiziert.

^dMU2 (Oldiges 2004): $D \approx 0 \text{ h}^{-1}$ (Fed-Batch), $BTM = 21, 31 \text{ g l}^{-1}$, Quenching: -40°C MeOH (60% v/v), Extraktion: -18°C HClO_4 (43,8% w/v) mit anschließender Neutralisation (gesättigte K_2CO_3 -Lösung), Analytische Methode: LC-MS-MS.

^e(Chassagnole et al. 2002): $D=(0, 1 \pm 0, 002) \text{ h}^{-1}$, $BTM = 8, 7 \text{ g DW l}^{-1}$, Quenching: flüssiger N_2 (G6P, F6P, PEP) bzw. -25°C HClO_4 (35% w/v) (FBP, ATP), Extraktion: 90°C Tris- H_2SO_4 /EDTA (G6P, F6P, PEP) bzw. -25°C HClO_4 (35% w/v) mit anschließender Neutralisation (2 M KOH/0,5 M Imidazol) (FBP, ATP), Analytische Methode: enzymatische Assays (G6P, F6P, FBP, PEP) bzw. HPLC (ATP).

^f(Theobald et al. 1997b): $D=(0, 1 \pm 0, 002) \text{ h}^{-1}$, $BTM = 15 \pm 0, 45 \text{ g DW l}^{-1}$; Quenching und Extraktion: -40°C HClO_4 (35% w/v) mit anschließender Neutralisation (2 M KOH/0,5 M Imidazol); Analytische Methode: enzymatische Assays (G6P, F6P, FBP, 3PG, PEP) bzw. HPLC (ATP).

^gKonzentration im Cytosol.

^h(Buziol et al. 2002): $D=(0, 1 \pm 0, 002) \text{ h}^{-1}$, $BTM \approx 15 \text{ g DW l}^{-1}$, Quenching: flüssiger N_2 , Extraktion: -25°C HClO_4 (35% w/v) mit anschließender Neutralisation (2 M KOH/0,5 M Imidazol), Analytische Methode: enzymatische Assays (G6P) bzw. HPLC (ATP).

ⁱ(Buziol et al. 2002): $D=(0, 1 \pm 0, 002) \text{ h}^{-1}$, $BTM \approx 10 \text{ g DW l}^{-1}$, Quenching: flüssiger N_2 , Extraktion: 90°C Tris- H_2SO_4 /EDTA, Analytische Methode: enzymatische Assays.

^j(van Dam et al. 2002): $D=0, 05 \text{ h}^{-1}$, $BTM = 14 \text{ g DW l}^{-1}$, Quenching: -40°C MeOH/ H_2O (40/60%), Extraktion: kochender EtOH mit anschließender Gefriertrocknung, Analytische Methode: LC-MS-MS.

^k(Visser et al. 2002): $D=0, 05 \text{ h}^{-1}$, $BTM = 14, 55 \text{ g DW l}^{-1}$, Quenching: -40°C MeOH/ H_2O (40/60%), Extraktion: kochender EtOH mit anschließender Gefriertrocknung, Analytische Methode: LC-MS-MS (G6P, F6P, FBP, 3PG, 2PG, PEP) bzw. enzymatische Assays (ATP).

^lDie intrazellulären Metabolitkonzentrationen wurden für ein spezifisches Zellvolumen $K_E \approx 27, 94 \text{ g DW} \cdot \text{l}^{-1}$ berechnet, das aus (van Dam et al. 2002, Tab. 6) für diesen Organismus abgeschätzt werden konnte.

^m(Lange et al. 2001): $D=0, 05 \text{ h}^{-1}$, $BTM \approx 15 \text{ g DW l}^{-1}$, Quenching: -40°C MeOH/ H_2O (40/60%), Extraktion: kochender EtOH mit anschließender Gefriertrocknung, Analytische Methode: LC-MS-MS (G6P, F6P, FBP, 3PG, 2PG, PEP) bzw. enzymatische Assays (ATP).

Bereits vor dem Glucose-Puls läßt sich in diesen Experimenten eine Erniedrigung der Konzentrationswerte einiger Metabolite feststellen, s. z.B. die Konzentration von G6P und F6P, PEP oder NAD in Anhang C, obwohl in diesem Zeitraum eigentlich ein Fließgleichgewichtszustand aufrechterhalten werden sollte. Dies läßt sich möglicherweise darauf zurückführen, daß bei Fed-Batch-Experimenten

aufgrund des fehlenden Abflusses von Fermentationslösung aus dem Reaktor generell nur ein sogenannter „Pseudo-Fließgleichgewichtszustand“ aufrechterhalten werden kann.

Sofern diese oder ähnliche Feststellungen zu der Entscheidung führten, bestimmte experimentelle Trajektorien nicht in die Modelle einzubeziehen, werden sie in Kapitel 5 im Rahmen der Erstellung konkreter Modelle erwähnt.

4.7 Zeitreihenanalyse

4.7.1 Einführung

Ziel der Zeitreihenanalyse nach (Arkin und Ross 1995) ist es, allein auf der Basis von zeitabhängigen Meßsignalen das zugehörige Reaktionsnetzwerk induktiv abzuleiten. Auf diese Weise kann das abgeleitete Reaktionsnetzwerk mit Vorkenntnissen über das System verglichen, und dadurch die Meßdaten auf evtl. auftretende Inkonsistenzen überprüft werden. Im Rahmen der Arbeit wurde die Eignung der Methode zur Datenvalidierung genauer untersucht.

Das Verfahren wird anhand der in Abb. 4–7 dargestellten linearen Reaktionssequenz erläutert. Über numerische Integration des DGL-Systems, s. Abb. 4–7 links, wurde ein *in-silico* Meßdatensatz erzeugt, s. Abb. 4–7 rechts. Er wurde der Zeitreihenanalyse (TSA) unterworfen, um das ursprünglich angenommene Reaktionsschema, s. Abb. 4–7 oben, herzuleiten.

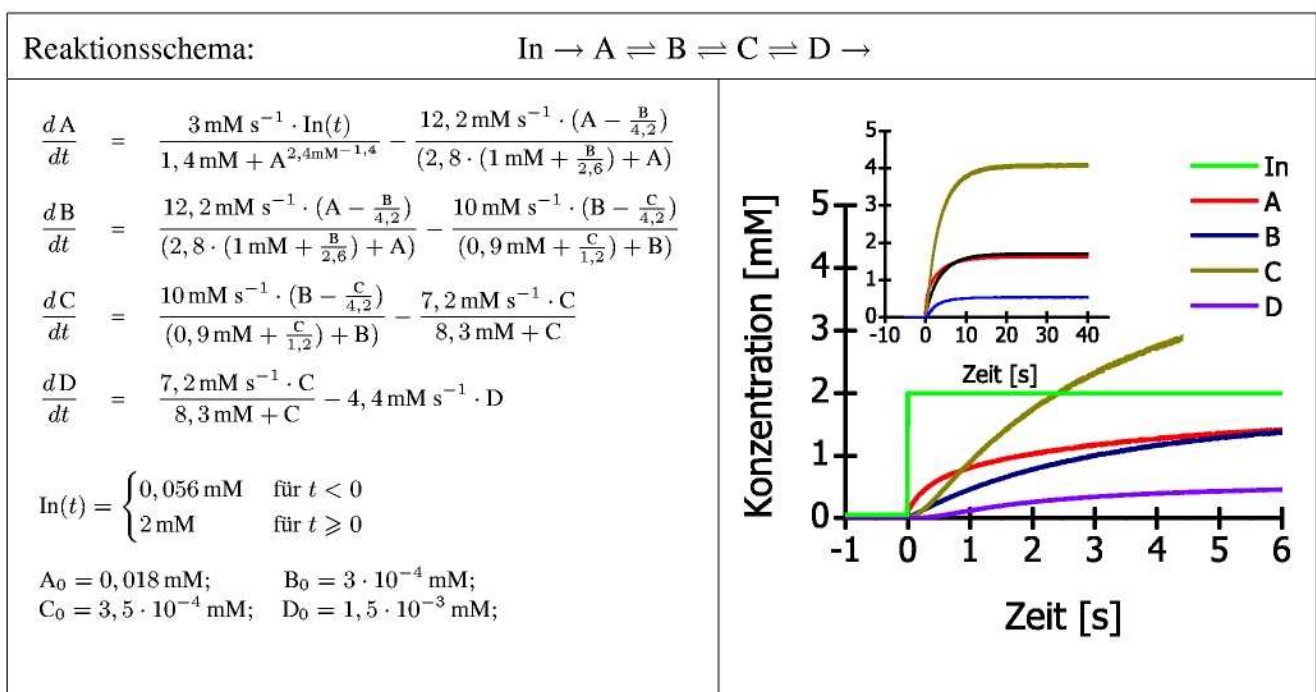


Abb. 4–7 Differentialgleichungssystem und zugehörige Ausgabefunktionen einer linearen Reaktionssequenz zur Verdeutlichung des TSA-Verfahrens

Grundlage des Verfahrens bilden die paarweisen Kreuzkorrelationsfunktionen, die den Ähnlichkeitsgrad zweier Trajektorien innerhalb gegeneinander versetzter Zeitsegmente wiedergeben. Sie werden aus den Kreuzkovarianzfunktionen $Cov_{jk}(\tau)$ berechnet:

$$Cov_{jk}(\tau) = \frac{1}{n} \sum_{t=\tau+1}^n (\tilde{c}_{j,(t-\tau)} - \bar{c}_j)(\tilde{c}_{k,t} - \bar{c}_k), \quad (4-5)$$

wobei die Zeitverschiebung $\tau = 0, 1, \dots, \nu$ angibt, um wieviel Zeiteinheiten die Abschnitte der Trajektorien gegeneinander versetzt sind. $j, k = 1, \dots, m$ sind die Indizes für die chemischen Spezies und $\bar{c}_j$ bzw. $\bar{c}_k$ der über alle n Zeitpunkte arithmetisch gemittelte Wert der Konzentration der Spezies j bzw. k . Die Kreuzkorrelationsfunktionen, $R_{jk}(\tau)$, ergeben sich dann aus Gleichung (4-5) wie folgt:

$$R_{jk}(\tau) = \frac{Cov_{jk}(\tau)}{\sqrt{Cov_{jj}(0)Cov_{kk}(0)}}, \quad -1 \leq R_{jk}(\tau) \leq 1. \quad (4-6)$$

In Abbildung 4-8 sind die Kreuzkorrelationsfunktionen zwischen dem Eingangssignal der Reaktionssequenz In und den übrigen Reaktanden gegen die Zeitverschiebung τ aufgetragen. Die Größe der Funktionsmaxima indiziert, wie eng die Verwandtschaft der Spezies ist. Die Lage der Maxima zeigt die zeitliche Abfolge, mit der das Eingabeprofil durch die Reaktionssequenz weitergeleitet wird, auf. Es ist

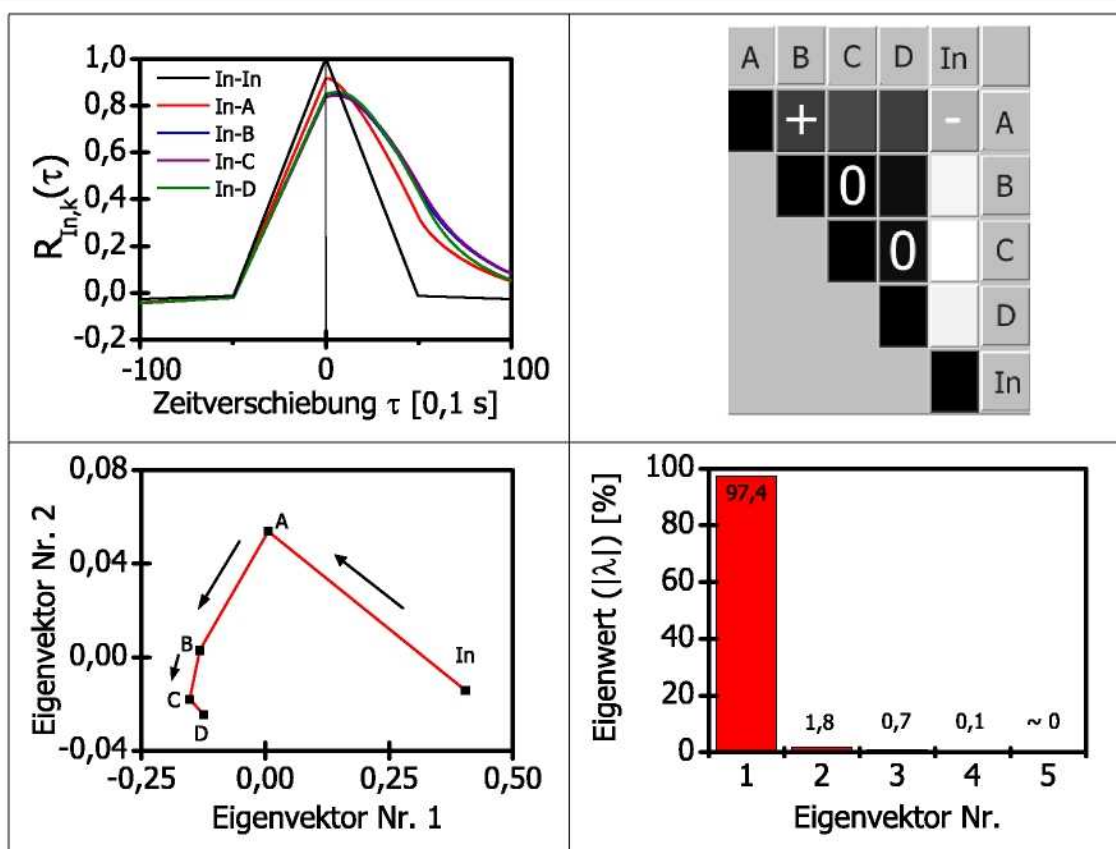


Abb. 4-8 Ergebnisse der Zeitreihenanalyse für das Beispiel in Abb. 4-7; links oben: Kreuzkorrelationsfunktionen des Eingangssignals, In, in Abhängigkeit von der Zeitverschiebung; rechts oben: Visualisierung der Distanzmatrix, die kleinste Distanz ($d_{B,C}=0,03$) bzw. die größte Distanz ($d_{In,A}=0,55$) entspricht dem RGB-Farbwert (0,0,0) (schwarz) bzw. (255,255,255) (weiß), die (+,-,0)-Einträge in den Matrixelementen bezeichnen, daß eine signifikante Korrelation zwischen den Spezies vorhanden ist und die Spezies in der Spalte vor, nach bzw. zeitgleich mit der Spezies in der Reihe auf das Eingangssignal reagiert; links unten: Eigenvektordiagramm, die Linien kennzeichnen signifikante Korrelationen, die Pfeile die zeitliche Abfolge der Reaktionen; rechts unten: Auftragung der Beträge der Eigenwerte für die zugehörigen Eigenvektoren

erwartungsgemäß erkennbar, daß das Maximum der Autokorrelationsfunktion für eine Zeitverschiebung $\tau = 0$ am größten ist ($\max |R_{In,In}|_{\tau} = 1$), denn hier wird die Trajektorie von In gerade mit sich selbst verglichen. Die Kreuzkorrelationsfunktion zwischen In und A weist etwas zeitlich versetzt, bei $\tau = 4$, ein kleineres Korrelationsmaximum $\max |R_{In,A}|_l = 0,92$ auf. Die Maxima der anderen Funktionen besitzen alle einen ungefähr ähnlichen Wert $\max |R_{In,k}|_l \approx 0,86, k = B, C, D$ und liegen bei einer

Zeitverschiebung von $\tau = 30$, $\tau = 50$ bzw. $\tau = 58$. Es läßt sich also bereits aus der Analyse der ersten 4 Kreuzkorrelationsfunktionen ableiten, daß das Eingangssignal In zunächst zu A und daraufhin A zu B reagiert.

Anstatt alle weiteren relevanten Kreuzkorrelationsfunktionen zu analysieren, was rasch unübersichtlich wird, wurde zur Kontraktion der Information und insbesondere zur graphischen Veranschaulichung eine Multidimensionale Skalierung (MDS)^{4.10} durchgeführt, s. z.B. (Mardia et al. 1979, S. 397 ff.). Dazu wurde die Matrix mit den Korrelationsmaxima zunächst in eine euklidische Distanzmatrix **D** transformiert:

$$\begin{aligned} d_{jk} &= (a_{jj} - 2a_{jk} + a_{kk})^{\frac{1}{2}} = 2^{\frac{1}{2}} \sqrt{1 - a_{jk}}, \\ a_{jk} &= \max |R_{jk}|_{\tau}, \end{aligned} \quad (4-7)$$

die den Grad der Unähnlichkeit zwischen zwei Spezies wiedergibt, s. in Abb. 4–8 (rechts oben) eine Visualisierung der Distanzmatrix für das Beispiel. Aus der Distanzmatrix wurde dann das zentrierte innere Matrixprodukt **B** berechnet:

$$\begin{aligned} \mathbf{B} &= -\frac{1}{2} \mathbf{H}(d_{jk}^2) \mathbf{H}, \\ \mathbf{H} &= \mathbf{I} - \frac{1}{m} \mathbf{1} \mathbf{1}^T. \end{aligned} \quad (4-8)$$

H bezeichnet die Zentrierungsmatrix, **I** die $m \times m$ Einheitsmatrix und $\mathbf{1} \mathbf{1}^T$ die $m \times m$ Identitätsmatrix. Zur metrischen Skalierung wurde die Matrix **B** einer Eigenvektoranalyse unterzogen, wobei $\|\gamma_j\|_2 = \lambda_j$ (λ_j : zu dem Eigenvektor γ_j gehöriger Eigenwert) gesetzt wurde, und die Eigenwerte abfallend, entsprechend ihrer Beträge sortiert wurden.

Die Resultate der MDS sind in Abb. 4–8 (unten) gezeigt. Anstelle der ursprünglichen Betrachtung von $(m^2 - m)/2$ Korrelationsmaxima in $\mathbb{R}^{m \times m}$ (m : Anzahl der Spezies) werden nur mehr m Punkte in eine durch die ersten beiden Eigenvektoren aufgespannte Ebene aufgetragen, s. Abb. 4–8 (links unten). Der Informationsverlust infolge der MDS wird durch die Analyse der Eigenwerte wiedergegeben: bei dem gezeigten Beispielsystem bleiben 99,2%=(97,4+1,8)% der Information über das Gesamtojekt durch die Projektion auf die ersten beiden Eigenvektoren erhalten, s. Abb. 4–8 (rechts unten).

Das Eigenvektordiagramm in Abb. 4–8 ist eine Visualisierung des Reaktionsschemas, das durch die TSA aus den Meßdaten abgeleitet wurde. Je dichter die Punkte im Diagramm beieinander liegen, desto größer ist ihre Verwandtschaft. Zusätzlich kennzeichnen die verbindenden Linien signifikante Korrelationen zwischen den betreffenden Spezies, die auf der Basis einer hierarchischen Clusteranalyse der globalen Korrelationsmaxima identifiziert wurden, s. folgenden Abschnitt^{4.11}. Die Pfeile resultieren aus der Analyse der jeweiligen Zeitverschiebungen der Korrelationsmaxima und geben die zeitliche Reihenfolge wieder, mit der die Spezies auf das Eingangssignal reagieren.

Im Fall des in Abb. 4–7 gezeigten Beispielsystems kann also das zur Erzeugung der Meßdaten verwendete Reaktionsschema wieder aus den Daten hergeleitet werden.

4.7.2 Implementierung der Zeitreihenanalyse

Zur automatisierten Durchführung des Verfahrens wurden C-Routinen für die einzelnen Arbeitsschritte implementiert, die unabhängig voneinander ausgeführt werden können, s. das Flußschema in Abb. 4–9.

^{4.10}Die Multidimensionale Skalierung ist ein Verfahren der multivariaten Statistik, durch das die Objekte einer Gesamtheit entsprechend ihrer Affinität in einem möglichst nieder-dimensionalen Raum gruppiert werden. Objekte mit einer großen Ähnlichkeit werden dabei durch eine besonders geringe Entfernung gekennzeichnet.

^{4.11}Signifikante Korrelationen entsprechen i. a. chemischen Reaktionen.

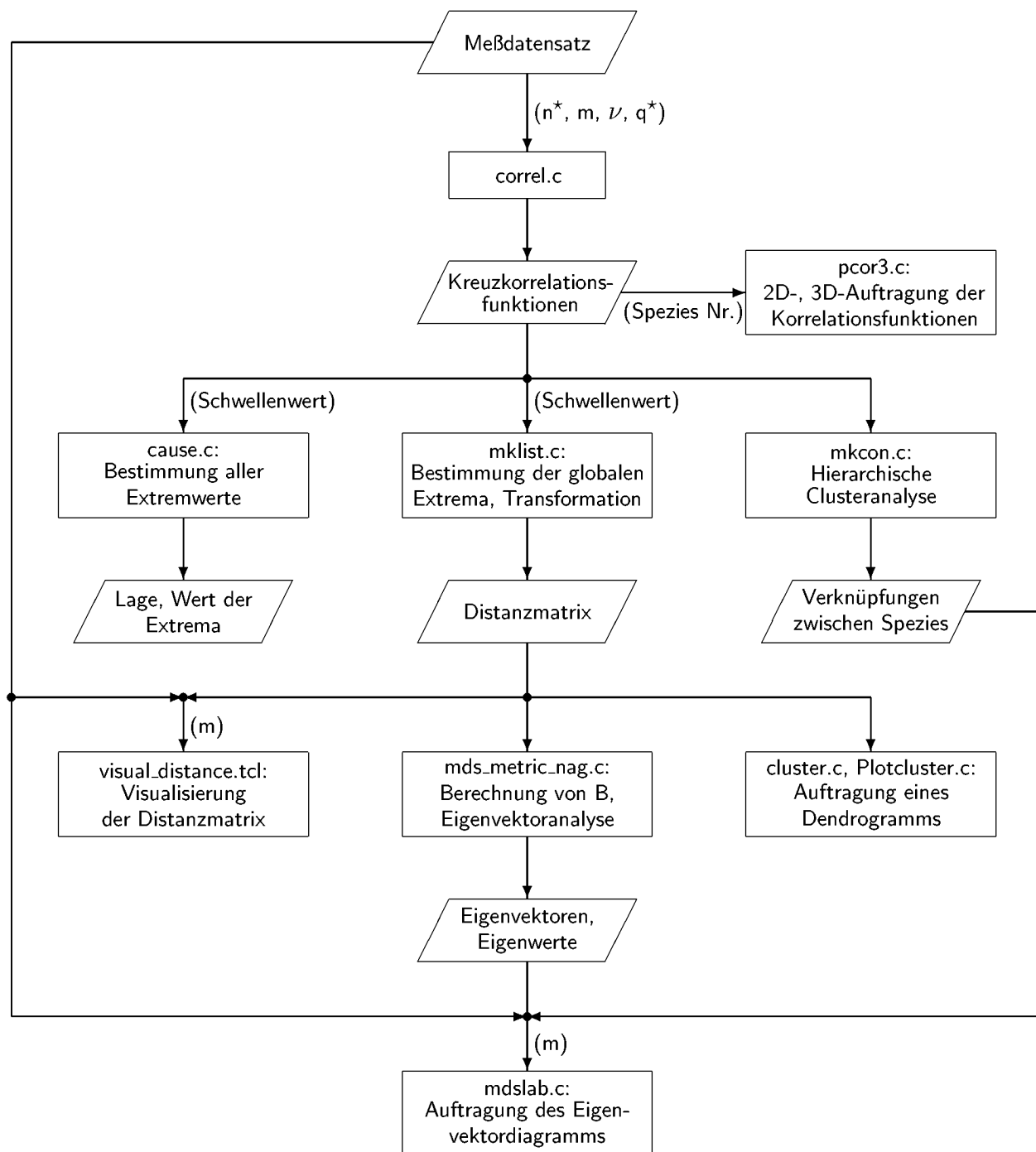


Abb. 4–9 Flußdiagramm der TSA-Implementation^{4.12}; Übergabeparameter: q^* , n^* : Parameter, die die Anzahl der zu betrachtenden Meßpunkte bestimmen; Schwellenwert: untere Grenze für den absoluten Wert des Extremums

Für eine vereinfachte Auswertung werden Diagramme erzeugt und als Postscript-Grafiken exportiert, wozu das Zeichenprogramm Gnuplot^{4.13} von den Routinen „pcor3.c“ bzw. „mdslab.c“ aufgerufen wird.

In der Routine „mkcon.c“ wurde ein Algorithmus für eine agglomerative, hierarchische Clusteranalyse implementiert, der die in den Eigenvektordiagrammen gezeigten Linien erzeugt und hier näher erläutert werden soll, s. Abb. 4–10. Nach dem Aufruf der Routine wird zunächst während der Initialisierung jede Spezies in eine eigene Gruppe eingeteilt. Iterativ wird dann jede Spezies aus einer Gruppe mit allen Spezies aus allen anderen aktuell vorhandenen Gruppen verglichen und das insgesamt betragsmäßig

^{4.12}Routinen adaptiert aus (Arkin 2001b).

^{4.13}Gnuplot Version 3,7, Copyright 1986-1993,1998,1999 Williams, T., Kellney, C., <http://www.ucc.ie/gnuplot/gnuplot-faq.htm>.

größte Korrelationsextremum bestimmt. Infolgedessen werden die beiden Spezies verknüpft, die Gruppen zusammengeschlossen und die Zahl der aktuell vorhandenen Gruppen um eins erniedrigt. Dementsprechend läuft der Algorithmus solange, bis nur noch eine Gruppe vorhanden ist.

Auf eine umfassendere Darstellung der weiteren Routinen wird verzichtet. Es sei nur noch darauf hingewiesen, daß zur Eigenvektoranalyse die Routine „mds_metric_nag.c“ die Funktion nagf02 der NAG C Bibliothek^{4,7} aufruft.

<pre> function MKCON(Kreuzkorrelationen) <i>Initialisierung</i> for i=1 to Spezies do Gruppe(i,0)←i; ZahlReihenInGruppe(i)←1; <i>Beginn der Gruppierung</i> do for i=1 to AktuelleGruppen rmax← (-2); for j=1 to AktuelleGruppen do if (j not equal to i) then (r,s1,s2)← FINDEMAXIMUM(i,j,Gruppe,ZahlReihenInGruppe); if (r greater than rmax) then rmax ← r; AktuelleGruppen← ASSOZIIEREGRUPPEN(i,j,Gruppe,ZahlReihenInGruppe); AUSGABE(r,s1,s2); while (AktuelleGruppen greater than 1); end; </pre>	<pre> function FINDEMAXIMUM(i,j,Gruppe,ZahlReihenInGruppe) returns (r,s1,s2) rmax← (-2); for alle ZahlReihenInGruppe (u) for alle ZahlReihenInGruppe (v) a←Gruppe(i,u) b←Gruppe(j,v) r ← max R_{a,b}(l) ; if (r greater than rmax) then rmax←r; s1←a; s2←b end; function ASSOZIIEREGRUPPEN(i,j,Gruppe,ZRInGruppe) returns (AktuelleGruppen) <i>Anschließen der Gruppe(j) an die Gruppe(i)</i> for alle ZahlReihenInGruppe (k) Gruppe(i, ZRInGruppe(i)+k)←Gruppe(j,k); ZRInGruppe(i)←ZRInGruppe(i)+ZRInGruppe(j); <i>Verwerfen der Gruppe(j)</i> for alle Gruppen mit höherem Index als j (k) Gruppe(k-1,...)←Gruppe(k,...); ZRInGruppe (k-1)=ZRInGruppe(k) AktuelleGruppen← AktuelleGruppen-1; end; </pre>
---	--

Abb. 4–10 Vereinfachter Algorithmus der in der Routine „mkcon“ durchgeführten hierarchischen Clusteranalyse

4.7.3 Test des Verfahrens anhand von *in-silico* Datensätzen

Da das Verfahren zur Datenvalidierung eingesetzt werden sollte, wurde auf der Basis von kleineren *in-silico* Meßdatensätzen geprüft, inwiefern zunehmend verrauschte bzw. numerisch manipulierte Trajektorien die Ergebnisse beeinflussen. Anhand von größeren Datensätzen wurde getestet, ob die Zeitreihenanalyse auch für komplex interagierende Systeme ein sinnvolles Reaktionsschema ableiten kann.

Zur Untersuchung des Einflusses verrauschter Trajektorien auf die Ergebnisse der TSA wurde allen Ausgabefunktionen der in Abb. 4–7 dargestellten linearen Reaktionssequenz additiv ein Rauschen w überlagert:

$$\begin{aligned}\tilde{c}_{i,j}^* &= \tilde{c}_{jk} + w_{i,j}, \\ w_{i,j} &\sim N(0, \sigma^2), \quad \sigma^2 = \left(\frac{f_{rel}}{100} |\tilde{c}_{i,j}| \right)^2,\end{aligned}\tag{4-9}$$

wobei $\tilde{c}_{i,j}$ bzw. $\tilde{c}_{i,j}^*$ Datenpunkte des Reaktanden j zum Zeitpunkt i aus dem originalen bzw. verrauschten *in-silico* Datensatz bezeichnen und $N(0, \sigma^2)$ für die Normalverteilung mit der Varianz σ^2 steht. Die Varianz des Rauschens wurde über den für alle Trajektorien gleichen Faktor f_{rel} variiert, und die

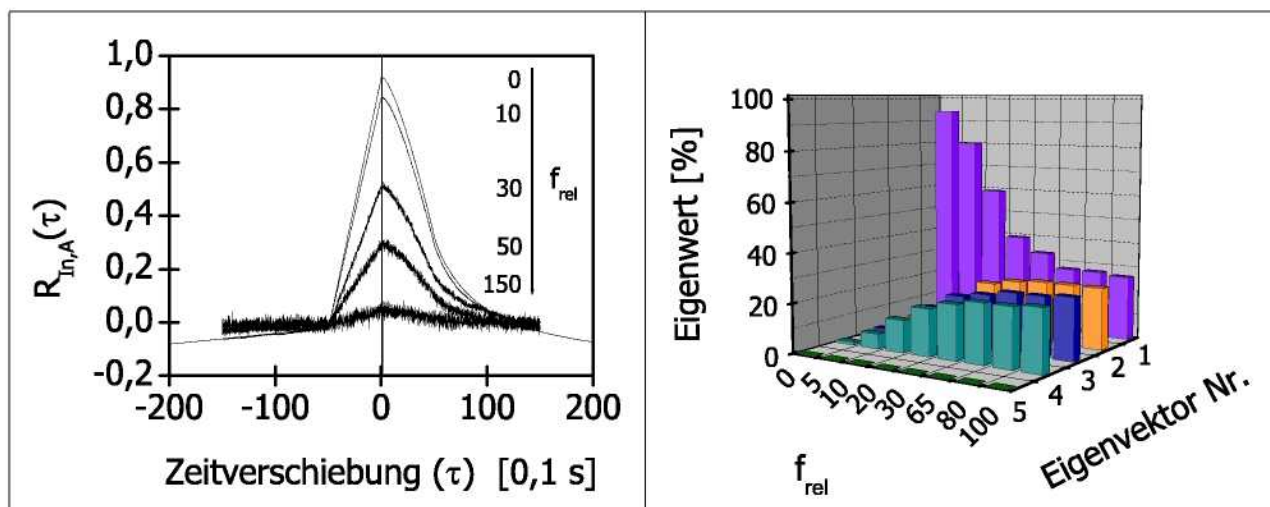


Abb. 4–11 Ergebnisse der TSA für verrauschte Ausgabefunktionen des Modells in Abb. 4–7, s. Gleichung (4–9); links: Kreuzkorrelationsfunktionen der Spezies In mit der Spezies A, rechts: aus der MDS resultierende Eigenwerte in Abhängigkeit von der Rauschstärke f_{rel} ($f_{rel}=0$ kennzeichnet den originalen Meßdatensatz)

resultierende Schar der Kreuzkorrelationsfunktionen sowie die Verteilung der Eigenwerte aufgetragen.

Aus Abb. 4–11 ist ersichtlich, daß für höhere Werte von f_{rel} die Größe der Korrelationsmaxima abnimmt, während ihre Lage unverändert bei $\tau \approx 0,00$ s bleibt. Die Eigenwerte 1-4 gleichen sich mit zunehmender Varianz des Rauschens aneinander an und liegen ab einem Wert von $f_{rel} \geq 30$ bei ca. 25%. Dies entspricht einer völligen statistischen Unabhängigkeit der Trajektorien der Spezies, A, B, C, D^{4,14}. Die hier nicht gezeigten Eigenvektordiagramme, d.h. die Projektionen des Gesamtobjektes auf die beiden wichtigsten Eigenvektoren, lassen sich ab einem Wert von $f_{rel} \approx 15 - 20$ nicht mehr eindeutig interpretieren.

Um die Auswirkung inkonsistent verlaufender Trajektorien zu untersuchen, wurde die aus dem Modell in Abb. 4–7 resultierende Ausgabefunktion von A ab dem Zeitpunkt $t_0 = 0$ s durch numerische Funktionen in Abhängigkeit von der Zeit ersetzt, s. Tab. 4–4. Die Ergebnisse der Zeitreihenanalyse sind in Abb. 4–12 zusammengefaßt.

Während Lage und Wert der Korrelationsmaxima für die manipulierten Funktionen M3 und M4 ungefähr erhalten bleiben, weil diese als hyperbolische Kurven dem ursprünglichen Verlauf von A ähneln, werden die Kreuzkorrelationsfunktionen der anderen Funktionen deutlich deformiert. Dabei besitzen die für die Manipulationen M1 und M2 resultierenden Korrelationsfunktionen keinen Peak-förmigen, sondern einen eher flachen Verlauf mit einem geringen Maximum $R_{In,A}(\tau) \approx 0,5$ bei einer großen zeitlichen Verschiebung $\tau = 9,5$ s. Bei dieser Zeitverschiebung wird offensichtlich gerade der sigmoide Anstieg numerischen Funktion zum Zeitpunkt $t \approx 10$ s mit dem Anstieg des Eingangssignals In zum Zeitpunkt $t = 0$ s verglichen. Die Korrelationsfunktionen aus den Manipulationen M5 und M6 weisen zwei kleinere Extrema auf. Die Maxima bei $\tau \approx 0$ s resultieren aus dem Vergleich der ansteigenden Abschnitte der Funktionen M5 und M6 zum Zeitpunkt $t = 0$ s mit dem Anstieg des Eingangssignals bei $t \approx 0$ s. Die Minima bei $\tau \approx 5$ s (M5) bzw. $\tau \approx 10$ s (M6) ergeben sich offensichtlich aus dem Vergleich der zur Zeit $t = 5$ s (M5) bzw. $t = 10$ s (M6) abfallenden Abschnitte der numerischen Funktion mit dem ansteigenden Abschnitt des Eingangssignals zum Zeitpunkt $t = 0$ s. Nach der Eigenvektoranalyse der Distanzmatrix tragen bei allen manipulierten Datensätzen jeweils zwei Eigenvektoren wesentlich zur Wiedergabe des Gesamtobjektes bei, während die Eigenwerte der anderen Eigenvektoren vernachlässigbar gering sind, s. Abb. 4–12. Dies stellt einen deutlichen Unterschied zu den Ergebnissen für die TSA

^{4,14}Die Trajektorie von In wurde nicht verrauscht.

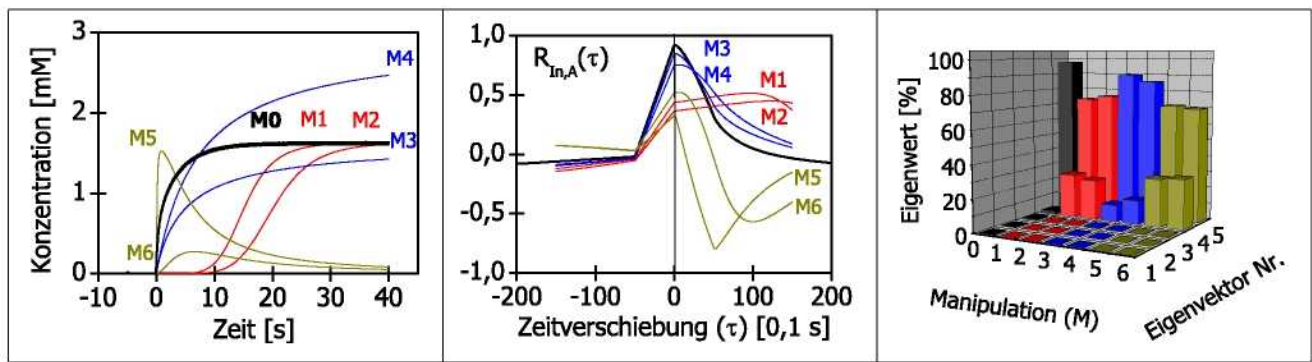


Abb. 4–12 Ergebnisse der TSA für inkonsistent verlaufende Trajektorien der Spezies A; links: numerisch manipulierte Trajektorie der Spezies A bei ansonsten unverändertem Modell aus Abb. 4–7; Mitte: resultierende Kreuzkorrelationsfunktion der Spezies In und A; rechts: Eigenwerte der MDS (Manipulation M0 bzw. dickere Linien kennzeichnen den originalen Meßdatensatz)

Tab. 4–4 Numerische Funktionen zur Ersetzung der Ausgabefunktion A aus dem Modell in Abb. 4–7 ab einem Zeitpunkt von 0 s

Funktion	Parameter	M1	M2	M3	M4	M5	M6	Einheit
$\frac{P_1^n \cdot t^n}{t^n + P_2^n}$	P_1	1,087	1,087	1,709	2,909	–	–	$[\text{mM}]^{1-n}$
	n	5,932	4,932	0,923	1,000	–	–	[1]
	P_2	15,127	20,127	5,127	7,127	–	–	[s]
$\frac{P_1^n \cdot t^n}{(t^2/P_3 + P_2)^n + t^n}$	P_1	–	–	–	–	1,409	1,409	$[\text{mM}]^{1-n}$
	n	–	–	–	–	1,600	1,600	[1]
	P_2	–	–	–	–	0,127	9,327	[s]
	P_3	–	–	–	–	6,200	4,620	[s]

der verrauschten Trajektorien dar, bei der mit wachsendem Rauschanteil keine dominierenden Eigenvektoren mehr identifiziert werden können, da sich die Abstände zwischen den einzelnen Eigenwerten nivellieren.

Der Einfluß verrauschter bzw. teilweise inkonsistenter Trajektorien wurde anhand von weiteren vier *in-silico* Datensätzen unterschiedlicher Größe analysiert, wobei die folgende Vorgehensweise zur Identifikation der Störeinflüsse abgeleitet wurde:

1. Zunächst sollten die Resultate aus der MDS untersucht werden:

- wird eine hohe Zahl von Eigenvektoren (N) in Bezug auf die Zahl der betrachteten Spezies benötigt ($N > 40\%$) um über 90% des Gesamtobjektes wiederzugeben, zeigt dies insgesamt nur schwache Korrelationen zwischen den Meßdaten an, deren Ursache beide Störeinflüsse, d.h. Rauschen und inkonsistente Trajektorien, sein können;
- besitzen in einem solchen Fall nur wenige Eigenwerte einen Wert kleiner als (1-2)% und unterscheiden sich die Beträge zweier aufeinanderfolgender Eigenwerte an keiner Stelle deutlich voneinander, so trägt vermutlich mehr das Rauschen zu der geringen Korrelation der Daten bei;
- ist dies nicht der Fall, scheinen vorwiegend systematisch vom wahren Verlauf abweichende Trajektorien die hohe Dimensionalität des Objektes zu verursachen;
- das Eigenvektordiagramm sollte dahingehend geprüft werden, ob sich ein sinnvolles Reaktionsschema ableiten läßt, und ob evtl. nur die Lage einzelner Spezies im Diagramm nicht

erklärbar ist; im Zusammenhang mit dem letzten Punkt kann dies zur Identifikation etwaiger inkonsistenter Trajektorien beitragen;

2. Im Anschluß sollten die Kreuzkorrelationsfunktionen geprüft werden:

- Funktionen mit breiten, deformierten Peaks und geringen Extrema ($|R| \approx [0, 4 - 0, 6]$) weisen entweder auf inkonsistente Trajektorien oder auf in höherem Ausmaß inhibierte bzw. aktivierte Reaktionen der betreffenden Spezies hin;
- bei dem Vorliegen solcher Ergebnisse sollten die Kreuzkorrelationsfunktionen der betreffenden Spezies mit bekannten Effektoren auf evtl. auftretende starke Korrelationen untersucht werden;
- die zeitliche Verschiebung der Korrelationsextrema in einer Reaktionssequenz aufeinanderfolgender Spezies sollte immer $\tau \geq 0$ s sein, ansonsten deutet dies ebenfalls auf die Inkonsistenz bestimmter Trajektorien zu dem übrigen Datensatz hin;

3. Trajektorien, die infolge von Punkt 2 identifiziert wurden, sollten aus dem Datensatz entfernt, und die TSA erneut mit dem reduzierten Datensatz durchgeführt werden; ergibt sich eine Verbesserung der Resultate, so kann mit höherer Sicherheit angenommen werden, daß diese Trajektorien wenigstens zeitweise einen inkorrekten Verlauf aufweisen;

Infolge der Untersuchungen wurde weiterhin vorwiegend bei Systemen, die starken effektorischen Einflüssen unterliegen, festgestellt, daß das Potential der TSA zur Interpretation von Daten aus speziellen Reaktionssystemen limitiert ist, weil zur Ableitung des Reaktionsnetzwerkes lediglich die Ähnlichkeit der zeitabhängigen Signale herangezogen wird. Dies kann anhand von Abb. 4–12 demonstriert werden. Bei Vorliegen einer sehr stark inhibierten oder sehr langsamen Reaktion, $B \rightleftharpoons C$, in der linearen Reaktionssequenz aus Abb. 4–7 könnte nämlich ein ähnlicher Verlauf für die Spezies A erwartet werden, wie er durch die Manipulationen M5 bzw. M6 numerisch erzwungen wurde. Für ein solches Reaktionssystem würde dann die TSA keine eindeutigen Rückschlüsse auf den zugrundeliegenden Reaktionsmechanismus zulassen, s. Abb. 4–12.

Um diesbezüglich zu prüfen, ob die TSA zur Auswertung der sehr komplexen Datensätze aus den Substrat-Puls-Experimenten genügend Aussagekraft besitzt, wurde das in (Chassagnole et al. 2002) dargestellte Modell simuliert. Dieses Modell beschreibt das Verhalten von 17 intrazellulären Metaboliten der Glykolyse und des PPP während eines Substrat-Puls-Experimentes. Die Metabolite sind über 30 Reaktionen untereinander hoch vernetzt und sind dabei dem Einfluß von 13 verschiedenen Effektoren ausgesetzt. Somit weist das Modell eine ähnliche Komplexität auf, wie sie auch von den dieser Arbeit zugrundeliegenden Datensätzen erwartet wurde, so daß es für einen realistischen Test des Verfahrens besonders geeignet schien.

Die simulierten Ausgabefunktionen wurden also der TSA unterworfen, wobei die wichtigsten Ergebnisse in Abb. 4–13 dargestellt sind. Die Kreuzkorrelationsfunktionen der betrachteten Spezies G6P und PEP sind im Wesentlichen erklärbar. Das größte Maximum $R_{jj}(0 \text{ s}) = 1$ weisen die beiden Autokorrelationsfunktionen auf. Dem Maximum der Korrelationsfunktion zwischen G6P und G6P folgen die Extrema der Funktionen von F6P, G1P, 6PG und Ri5P mit G6P zeitlich nacheinander^{4.15}, was in Übereinstimmung mit dem im Modell angenommenen Reaktionsschema steht. Denn es wurde für das Modell festgelegt, daß die Reaktion zwischen G6P und F6P sehr schnell abläuft, G6P zu G1P und 6PG reagiert und daß in einer Folgereaktion aus 6PG Ri5P entsteht^{4.16}. Auf eine ähnliche Weise lassen sich die Korrelationsfunktionen von PEP mit den gezeigten Spezies erklären. Die Reaktion

^{4.15} G1P: Glucose-1-phosphat; 6PG: 6-Phosphogluconat; Ri5P: Ribulose-5-phosphat.

^{4.16} Die Reaktion von G6P zu 6PG wird durch NADPH inhibiert und ist somit langsamer als die von G6P zu G1P.

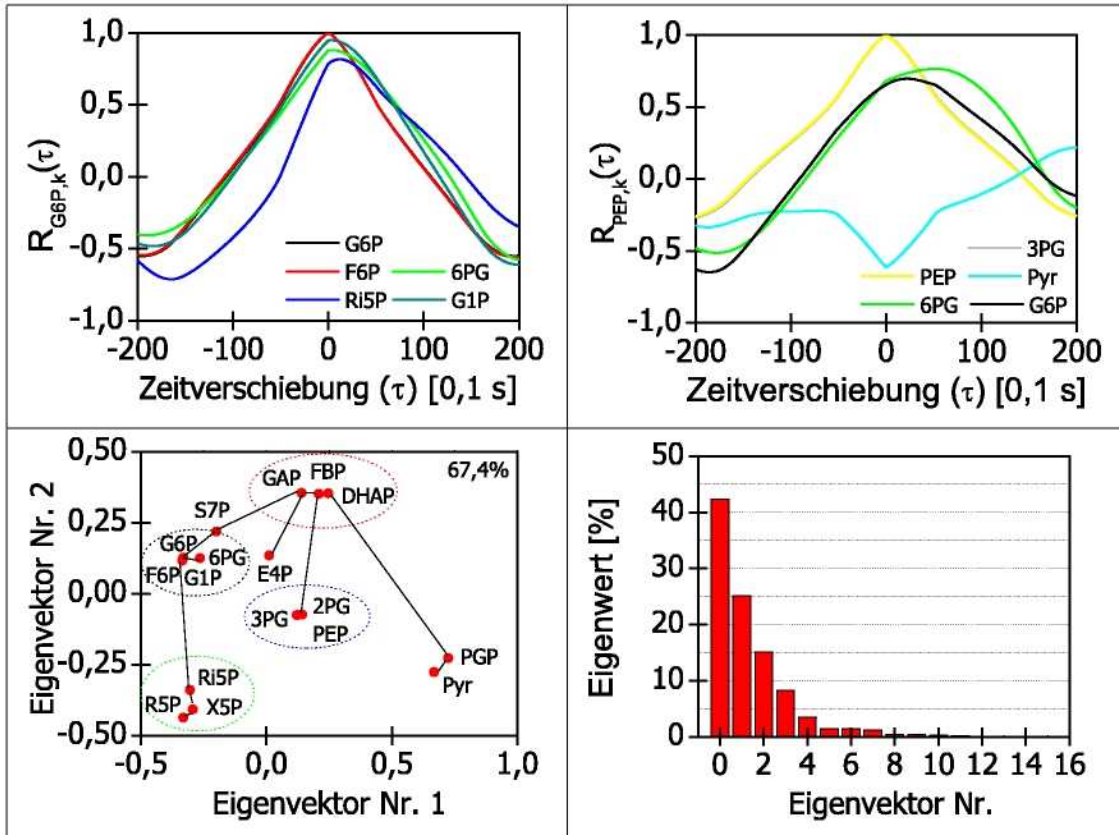


Abb. 4–13 Ergebnisse der TSA für die Ausgabefunktionen des Modells aus (Chassagnole et al. 2002); oben links: Kreuzkorrelationsfunktionen der Trajektorie von G6P, bzw. oben rechts: der Trajektorie von PEP mit Trajektorien anderer Spezies; unten links: Eigenvektordiagramm; unten rechts: Verteilung der Beträge der Eigenwerte

zwischen 3PG und PEP erfolgt wiederum so schnell, daß beide Korrelationsfunktionen aufeinander liegen, die Maxima von G6P und 6PG befinden sich bei einer positiven Zeitverschiebung τ , die Spezies reagieren also auf das PEP-Signal vermutlich infolge der Glucose-Aufnahme- und der darauffolgenden G6P-6PG-Reaktion. Die Korrelationsmaxima der Spezies sind jedoch weniger stark ausgeprägt: $R_{PEP,G6P}(17\text{ s}) \approx 0,7$ bzw. $R_{PEP,6PG}(45\text{ s}) \approx 0,76$. Dies weist darauf hin, daß der Zusammenhang zwischen PEP und G6P schwieriger zu deuten ist, weil zum einen G6P aus PEP bei der Glucose-Aufnahme gebildet wird und zum anderen PEP aus G6P infolge der Glykolyse-Reaktionen entsteht. Bemerkenswert ist die Korrelationsfunktion von PEP und Pyr. Obwohl die Metabolite sowohl bei der Glucose-Aufnahme als auch in der Glykolyse direkt miteinander reagieren, bildet sich kein deutliches Maximum aus. Auch bei der Betrachtung der simulierten und der experimentellen Ergebnisse für Pyr, s. (Chassagnole et al. 2002), fallen im Vergleich deutliche Differenzen zwischen den Trajektorien auf. Es ist also möglich, daß der Verlauf der PEP-Pyr-Korrelationsfunktionen evtl. auf ein inkonsistentes Simulationsergebnis für Pyr hin weist.

Nach der MDS zeigt die Untersuchung der Eigenwerte hinsichtlich der oben aufgeführten Punkte einen hohen Korrelationsgrad zwischen den simulierten Trajektorien. Und das obwohl die Kreuzkorrelationsfunktionen zwischen einzelnen Spezies, wie z.B. zwischen PEP und Pyr, nicht unbedingt einfach zu deuten sind. Denn über 90% des Gesamtobjektes können bereits durch die ersten vier Eigenvektoren wiedergegeben werden, s. Abb. 4–13 unten rechts. 12 der insgesamt 17 Eigenwerte sind betragsmäßig kleiner als 2%. Ferner lassen sich deutlich größere Unterschiede zwischen den Beträgen der ersten fünf Eigenwerten als zwischen den darauffolgenden 12 Eigenwerten feststellen. Eine Auftragung der Komponenten der ersten beiden Eigenvektoren, s. Abb. 4–13 unten links, spiegelt 67,4% der Information

des Gesamtobjektes wieder. Dieses 2D-Diagramm läßt bereits eindeutige Rückschlüsse auf das zugrundeliegende Reaktionsnetzwerk des Modells zu. Eine Erhöhung der Dimension der Darstellung führt dagegen nur zu einer unwesentlichen weiteren Verbesserung der Resultate. Der Cluster der unmittelbar auf den Glucose-Puls reagierenden Spezies G6P, F6P G1P und 6PG ist nach unten hin mit einem Cluster bestehend aus den untereinander eng verwandten C5-Zuckern des PPP und nach oben hin mit einem Cluster der Glykolyse-Intermediate FBP, GAP und DHAP verknüpft. Offenkundig wird die ursprünglich angenommene Aufteilung des Glucose-Aufnahmeflusses ausgehend von G6P in die Glykolyse und den PPP durch das Eigenvektordiagramm richtig dargestellt. Die zwischen den korrespondierenden Clustern liegenden einzelnen Spezies S7P und E4P deuten vermutlich die enge Verwandtschaft dieser Intermediate mit dem glykolytischen Intermediaten GAP bzw. F6P infolge der Transaldolase- und Transketolase-Reaktion an. Mit dem Cluster aus FBP, GAP und DHAP konnte ein aus den Metaboliten 3PG, 2PG und PEP bestehender Cluster auf Basis des Konnektionsalgorithmus verknüpft werden. Dies läßt sich insofern erklären, als die drei Spezies aus GAP bzw. DAHP in dem unteren Abschnitt der Glykolyse entstehen. Allerdings reagiert GAP über 1,3-Diphosphoglycerat (DGP) zu 3PG und PEP reagiert direkt zu Pyr, so daß DGP und Pyr eigentlich in dem Cluster aus 3PG, 2PG und PEP zu erwarten gewesen wären. Daher wurden die Kreuzkorrelationsfunktionen von DPG und Pyr genauer untersucht. In Bezug auf den sich inkonsistent verhaltenden Metaboliten Pyr wurden die Aussagen bereits oben getroffen. Die Korrelationsfunktionen von DGP mit GAP, 3PG und 2PG weisen alle einen relativ scharfen Peak mit einem Extremum $|R_{DGP,k}|(0\text{ s}) = 0,63$ auf, was die chemische Verwandtschaft der Metabolite widerspiegelt. Allerdings besitzt die Korrelationsfunktion zwischen DGP und Pyr ein noch größeres Minimum $R_{DGP,Pyr}(0\text{ s}) = -0,95$, auf das die Nähe der betreffenden Punkte im Eigenvektordiagramm zurückzuführen ist. Daher wurde nur die Trajektorie von Pyr ausgeschlossen und die TSA erneut wiederholt, wobei sich die Ergebnisse für das Eigenvektordiagramm verbesserten.

Aufgrund der Resultate für das in (Chassagnole et al. 2002) dargestellte Modell konnte also gezeigt werden, daß die TSA das zugrundeliegende Reaktionsnetzwerk auch aus komplex integrierenden Systemen in groben Zügen ableiten kann.

4.7.4 Anwendung auf Datensätze aus Substrat-Puls-Experimenten

Da alle bisherigen Tests des Verfahrens auf der Basis von *in-silico* Datensätzen mit positiven Ergebnissen abgeschlossen wurden, sollte die TSA auch auf die vorliegenden realen *in-vivo* Datensätze, s. Anhang C, angewendet werden. Die Ergebnisse für die interpolierten Datensätze aus den Experimenten WT1 und MU1 sind exemplarisch in Abb. 4–14 zusammengefaßt^{4.17}. Die aus der MDS resultierenden Eigenwerte nehmen nahezu linear ab, wobei nur der erste für MU1 bzw. die ersten beiden Eigenwerte für WT1 etwas nach oben hin von einer angenommenen Gerade abweichen. Es lassen sich also keine allzu deutlichen Unterschiede zwischen den Beträgen der größeren und der kleineren Eigenwerte feststellen. Nur wenige Eigenwerte, nämlich zwei bei Experiment MU1 bzw. fünf bei Experiment WT1, sind vernachlässigbar gering ($\leq 2\%$). Insgesamt werden für den Datensatz aus WT1 bzw. MU1 neun bzw. acht Eigenwerte, d.h. ein Anteil von über 56% bzw. über 66% Eigenwerten bezogen auf die Zahl der untersuchten Spezies, benötigt, um über 90% des Gesamtobjektes darzustellen. Entsprechend der Schlußfolgerungen auf Seite 62 weisen die Resultate also in erster Linie auf ein relativ starkes, den Trajektorien überlagertes Rauschen hin, welches hauptsächlich für den niedrigen Korrelationsgrad der Trajektorien verantwortlich ist^{4.18}. Dementsprechend lassen sich die Eigenvektordiagramme, die nur 38,9% bzw. 39,7% der Information über das Gesamtobjektes wiedergeben, nicht mehr eindeutig interpretieren, s. Abb. 4–14 (rechts). Um das evtl. störende Rauschen zu eliminieren, wurden die Trajektorien

^{4.17}Die Ergebnisse für das Experiment MU2 sind mit den anderen Resultaten vergleichbar und werden daher dort nicht gezeigt.

^{4.18}Diese Ergebnisse stehen offenkundig im Einklang mit der in Abschnitt 4.2 vorgestellten Fehlerrechnung, die relativ hohe Meßunsicherheiten für die intrazellulären Metabolitkonzentrationen ergab.

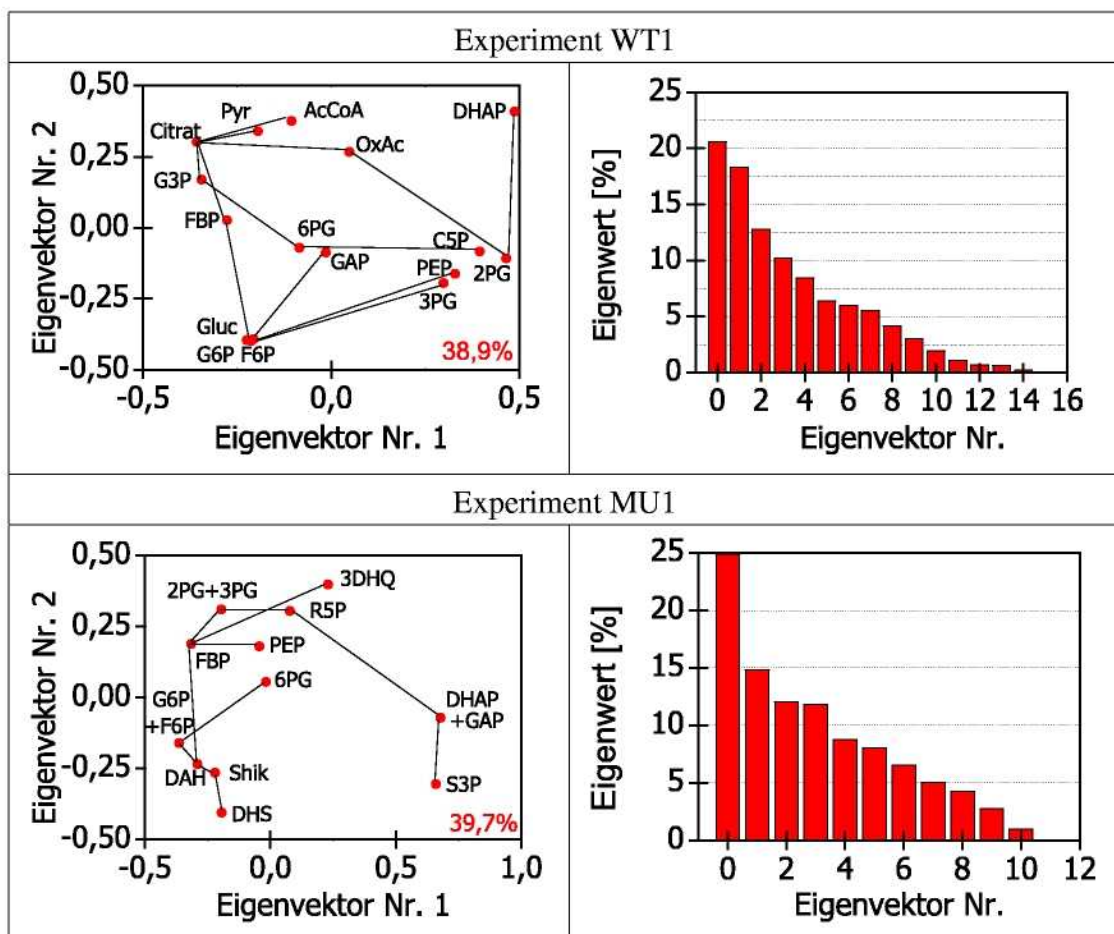


Abb. 4–14 Ergebnisse der TSA für die Datensätze aus den Experimenten WT1 (oben) und MU1 (unten); links befinden sich jeweils die Eigenvektordiagramme, rechts jeweils die Auftragungen der Eigenwerte

über die in Abschnitt 4.5 vorgestellten Methoden, insbesondere die FFT-Filterung, geglättet, bevor sie der TSA unterzogen wurden. Eine Verbesserung der Resultate ließ sich jedoch nicht erzielen. Dafür kann es verschiedene Gründe geben:

1. durch die vorgenommene Datenglättung wurde nicht nur das Rauschen entfernt, sondern die Trajektorien teilweise auch verfälscht;
2. neben dem Rauschen können sich inkonsistente Trajektorien als zusätzliche Störeinflüsse im originalen Datensatz befinden; wenn dies der Fall sein sollte, so müßten relativ viele Trajektorien dieser Art existieren, da die TSA auch mit zahlreichen, um verschiedene Trajektorien reduzierten, ungeglätteten Datensätzen ohne nennenswerte Verbesserungen durchgeführt wurde;
3. das TSA-Verfahren ist evtl. doch nicht robust genug, um das Reaktionsschema aus realen Daten abzuleiten. Denn es untersucht als einziges Kriterium für die chemische Verwandtschaft von Metaboliten die Ähnlichkeit zwischen den Trajektorien, s. die Diskussion auf S. 63 (Mitte). Weiterhin ist möglich, daß sich in einem komplexen Stoffwechselnetzwerk die Einflüsse vieler Spezies überlagern und die Interpretation der TSA-Resultate somit erschweren^{4,19};

Gegen diese Vermutungen sprechen allerdings die Resultate der TSA für das Modell aus (Chasnagnole et al. 2002), und die Beobachtungen aus anderen veröffentlichten Stoffwechselmodellen.

^{4,19} Beispielsweise könnten experimentell nicht bestimmte Metabolite einen signifikanten Einfluß auf die untersuchten Spezies ausüben oder sehr viele Effektoren regulativ auf einen Reaktionsschritt einwirken. Solche solchen Fällen würden dann allerdings auch die Bildung von DGL-Modellen beeinträchtigen, d.h. die DGL-Modelle müßten durch zusätzliche experimentelle Messungen abgesichert werden.

Bei den letzteren Modellen zeigt sich schon graphisch eine relativ große Ähnlichkeit zwischen den Trajektorien von Spezies, die miteinander reagieren oder sich effektorisch beeinflussen, s. z.B. (Theobald et al. 1997b; Vaseghi et al. 1999; Wolf et al. 2001; Visser et al. 2002);

Die Prüfung der Punkte 1 und 2 durch etwaige Wiederholungsexperimente war im Rahmen dieser Dissertation nicht durchführbar. Insofern wurde nur Punkt 3 genauer untersucht, s. folgenden Abschnitt.

4.7.5 Erweiterungen der Methode

Da die TSA keine zufriedenstellenden Ergebnisse bei der Auswertung der Datensätze aus den Experimenten WT1, MU1 und MU2 lieferte, wurden zur Verbesserung des Verfahrens weitere Punkte in Betracht gezogen, die im Folgenden dargestellt werden.

Ableitung des Reaktionsschemas allein aus der zeitlichen Verschiebung der Extrema anstelle der MDS^{4,20} Die in (Arkin und Ross 1995) beschriebene TSA berücksichtigt nur die globalen Korrelationsextrema und leitet aus ihnen über die MDS das Reaktionsschema ab. Dabei könnten neben den „echten“ evtl. sogenannte Scheinkorrelationen auftreten. Unter Scheinkorrelationen werden Korrelationsextrema zwischen Spezies verstanden, die nicht unmittelbar miteinander chemisch verwandt sind. Z.B. tritt bei der Analyse des Beispiels in Abb. 4–7 ein Maximum zwischen A und C auf, obwohl es keine Reaktion $A \rightarrow C$ gibt. Auf der anderen Seite werden lokale Extrema nicht analysiert. Dadurch kann die Information über eine direkte chemische Verwandtschaft in Spezialfällen verlorengehen, wie z.B. bei einem Reaktionssystem, bei dem A direkt zu C und über B zu C reagiert, wobei die Scheinkorrelation zwischen A und C aus der Reaktion $A \rightarrow B \rightarrow C$ größer ist als die echte aus der Reaktion $A \rightarrow B$. Daher wurde anhand von einfachen Beispielsystemen versucht, Regeln für die Rekonstruktion des Reaktionsnetzwerkes allein aus der Lage, d.h. der zeitlichen Verschiebung, der Extrema abzuleiten. Dabei wurde festgestellt, daß die Extrema von Schein- und echten Korrelationen nicht getrennt voneinander beobachtet werden können, sondern gemeinsam bei derselben zeitlichen Verschiebung auftreten. Weiterhin führen sehr langsame Reaktionsschritte innerhalb von linearen Sequenzen zu einer nahezu gleichen zeitlichen Verschiebung der Korrelationsmaxima von allen der langsamen Reaktion folgenden Spezies. Zusätzlich nimmt der Zeitabstand der Extrema direkt verwandter Spezies innerhalb einer linearen Sequenz ab.

Aus diesen Gründen konnten keine eindeutige Regeln für die Ableitung des Reaktionsschemas aus den Zeitverschiebungen identifiziert werden und das bisherige MDS-Verfahren erwies sich bei den betrachteten Beispielsysteme als ebenso gut, wenn nicht besser geeignet.

Dabei empfiehlt es sich, die Zeitverschiebungen aller Extrema, die über die Funktion „cause“ ausgegeben werden können, unterstützend zur Interpretation der Eigenvektordiagramme zu betrachten.

Verkürzung der zur Berechnung der Kreuzkorrelationsfunktionen betrachteten Zeitsegmente Im bisherigen Verfahren wurden Korrelationsfunktionen aus allen n Datenpunkten der beiden Trajektorien berechnet und demzufolge die Anzahl der betrachteten Datenpunkte bei der Berechnung der Korrelationsfunktionen in Abhängigkeit von einer jeweils um eins ansteigenden zeitlichen Verschiebung um eins verkürzt, s. Gleichung (4–5). Eine Schärfung etwaiger Korrelationspeaks ist über die geeignete Wahl eines Startzeitpunktes und der darauffolgenden Zahl der betrachteten Meßpunkte denkbar. Dadurch kann ein Zeitfenster festgelegt werden, innerhalb dessen sich die jeweilige Referenzspezies signifikant

^{4,20}Diese Untersuchung wurde aufgrund eines entsprechenden Hinweises von Prof. G. Dikta, FH Aachen, Abteilung Jülich am 19.12.2002 durchgeführt.

ändert, und die Korrelationen der anderen Spezies mit der Referenz innerhalb eines gleich großen versetzten Zeitfensters ermittelt werden. Für das Beispielsystem aus Abb. 4–7 bietet sich z.B. ein Zeitfenster von $0 \text{ s} \leq t \leq 15 \text{ s}$ an. Zu diesem Zweck wurde die Routine „correl.c“ so implementiert, daß sowohl der Parameter $q^* \geq 0$ als Startzeitpunkt als auch der Parameter $n^* \leq n$ als Zahl der zu betrachtenden Meßpunkte übergeben wird, s. Abb. 4–9, anstatt diese automatisch aus der Datendatei zu bestimmen. Mit dieser Vorgehensweise konnte bei einigen Beispielsystemen tatsächlich eine Schärfung der Korrelationspeaks erreicht werden. Es sollten jedoch nicht zu wenige Meßdatenpunkte und zu große Zeitverschiebungen ν gewählt werden, da dies leicht zum Auftreten von artifiziellen Peaks führt (als ungefähre Richtwerte erweisen sich: $n^* \geq n/4, \nu \leq n^*/2$).

Veränderung der Dimensionalität des Eigenvektordiagramms Die Projektion des Objektes auf die beiden wichtigsten Eigenvektoren, die bisher unabhängig vom betrachteten System in jedem Fall erfolgte, ist eigentlich theoretisch unbegründet. Denn bei stark korrelierten Meßdaten, s. das Beispiel in Abb. 4–7, ist bereits der Anteil des ersten Eigenvektors dominant, so daß eine 1D-Abbildung für eine geeignete Repräsentation des Gesamtsystems ausreicht. Sind die Meßdaten dagegen schwächer korreliert, werden evtl. mehr als zwei Eigenvektoren benötigt, um eine geeignete Darstellung zu erreichen. Deswegen wurde die in (Hartung und Elpelt 1995, Abschnitt 8.1.3) vorgestellte Methode genutzt, um die Dimension der Projektion graphisch zu bestimmen. Dazu werden die Punkte im Eigenwertdiagramm (s. z.B. Abb. 4–7, rechts unten) mit Geraden verbunden und die Zahl der wichtigsten Eigenvektoren als der Wert festgelegt, für den alle links liegenden Punkte durch steilere Geraden verbunden sind als die rechts liegenden Punkte. Ergab sich daraufhin, daß das Objekt auf mehr als 3 Eigenvektoren dargestellt werden sollte, wurden entsprechende Ebenenschnitte des mehrdimensionalen Objektes untersucht.

Zusammenfassung der Ergebnisse für die Erweiterungen des Verfahrens Die Ergebnisse der TSA für *in-silico* Datensätze konnten in Einzelfällen durch Berücksichtigung der Punkte 2 und 3 verbessert werden. Für die Datensätze aus den Experimenten WT1, MU1 und MU2 waren jedoch keine nennenswerten Veränderungen feststellbar. Es läßt sich also nicht sicher klären, welche der auf Seite 66 zusammengefaßten Punkte die Ursache für die Resultate der TSA bei diesen Meßdaten sind.

5 Erstellung und Validierung des konzeptuellen Modells

5.1 Einführung

In diesem Kapitel wird die Vorgehensweise zur Modellerstellung anhand von einigen DGL-Modellen erläutert, die im Rahmen dieser Doktorarbeit entwickelt wurden. In der Regel wurden zunächst einfache Modelle mit einer geringen Zahl an DGL bzw. einfachen Enzymkinetiken konstruiert und dann über die Hinzunahme weiterer Details schrittweise verfeinert. Ein solcher sogenannter „Top-Down-Approach“ wird in Abschnitt 5.2 auf der Basis von fünf Modellen dargestellt, die die Synthese aromatischer Aminosäuren in zwei *E.coli*-Mutanten beschreiben. Über 100 weitere Modelle wurden zur Wiedergabe der in Experiment WT1 bestimmten Intermediate aus der Glykolyse und dem PPP eines *E.coli*-Wildstamms generiert, s. (Degenring und Takors 2002; Degenring und Takors 2003). Eines dieser Modelle, das eine für diese Arbeit vergleichsweise hohe Komplexität besitzt, wird exemplarisch in Abschnitt 5.3 dokumentiert, die anderen Modelle sind in Anhang G zusammengefaßt.

5.2 Modelle der Aromatenbiosynthese in *E.coli* - Mutanten

5.2.1 Allgemeine Hypothesen und vereinfachende Annahmen

Die den folgenden Modellen zugrundeliegende Zielstellung ist die Beschreibung des Aromatenbiosynthese-Weges (AAA) in zwei *E.coli*-Mutanten mit einem für beide Organismen vergleichbaren Modell. In Anbetracht der Datensätze, s. Anhang C, und der Durchführung der Experimente MU1 und MU2, s. Abschnitt 3.2, sind einige Besonderheiten auffällig, die Art und Umfang der DGL-Modelle bestimmen. Sie werden im Folgenden aufgelistet:

1. Extrazelluläre Metabolite konnten nicht quantifiziert werden:

Infolgedessen können keine quantitativen Aussagen über das nach dem Substrat-Puls real im Reaktor anliegende Glucose-Signal (Experiment MU1) sowie über Konzentrationen von Intermediaten des AAA im Probenüberstand (Experimente MU1 und MU2) getroffen werden.

Daher ist es notwendig, bzgl. des Glucose-Profiles spezielle Annahmen für die Modellierung zu machen. Für die AAA-Intermediate ist aus (Gerigk 2001) bekannt, daß bei Verwendung ähnlicher Stämme wenigstens die Metabolite L-Phe, Shik und u.U. DAH bzw. DHS verstärkt im Fermentationsüberstand auftreten^{5.1}. Gleichzeitig geht aus den Meßdaten hervor, daß im AAA keine signifikanten Änderungen der intrazellulären Metabolitkonzentrationen infolge des Glucose-Pulses auftreten, d.h. daß evtl. ein höherer Fluß in den AAA nach dem Puls durch einen größeren Abfluß einiger Metabolite aus der Zelle kompensiert werden könnte. Daher ist für die Modellierung zu entscheiden, in welchem Ausmaß und von welchen AAA-Intermediaten ein solcher Abfluß angenommen werden darf.

2. Die intrazelluläre Konzentration des Metaboliten Pyruvat konnte nicht bestimmt werden:

Pyr spielt eine wichtige Rolle für die Glucose-Aufnahmereaktion, da es als Produkt direkt beteiligt ist und über die Reversibilität der PykA- bzw. PykF-Reaktionen u.U. auch das die Glucose-Aufnahme limitierende Substrat PEP bereitstellen kann. Infolgedessen ist es wiederum nötig, besondere Annahmen bzgl. der Glucose-Aufnahme zu machen.

^{5.1} Dies betrifft den in MU1 verwendeten *E.coli* 4pF20-Stamm (DAH, Shik, L-Phe) und die Stämme *E.coli* 4pF26 (DAH, L-Phe), *E.coli* 4pF69 (DAH (gering), DHS, Shik, L-Phe), *E.coli* 4pF81 (Shik, L-Phe). *E.coli* 4pF81 ist bis auf das in das Plasmid integrierte *aroF*^{WT} Gen mit dem in MU2 verwendeten Stamm *E.coli* 4pF78 genetisch identisch.

3. Das Verhalten der experimentell bestimmten [G6P+F6P]-Summenkonzentration kann phänomenologisch schwer interpretiert werden:

Denn es ist eigentlich bekannt (Ferenci 1996; Glaser und Brown 1955; Westwood und Doelle 1974; Schreyer und Bock 1980), daß sowohl die Glucose-Aufnahme bzw. G6P-Synthese als auch die beiden G6P hauptsächlich verbrauchenden Reaktionen (die durch die G6pdh bzw. die Pgi katalysierten Reaktionen) sehr schnell verlaufen. Da weiterhin nach dem Puls ein Glucose-Überschuß vorlag, und eine PEP-Limitierung experimentell nicht beobachtet wurde, wäre somit ein schneller Konzentrationsanstieg der Metabolite unmittelbar nach dem Puls und daraufhin eine rasche Einstellung eines Fließgleichgewichtszustands von G6P+F6P zu erwarten gewesen. Sofern also nur gängige biochemische Informationen zugrundegelegt werden, läßt sich der beobachtete, zeitlich versetzte Anstieg von G6P+F6P in MU1 und MU2 mit einem DGL-Modell vermutlich kaum modellieren.

4. Es wurden hochfrequente Schwankungen der Konzentrationswerte in dem Verlauf der Trajektorien beobachtet;

diesbezügliche Konsequenzen wurden bereits in den Abschnitten 3.2.3 und 4.5 abgeleitet;

5. Die Konzentrationen einiger Metabolite, wie z.B. NAD, G6P+F6P, nehmen im Zeitraum vor dem Substrat-Puls z.T. ab:

In diesem Zeitraum sollte jedoch eigentlich ein Fließgleichgewichtszustand aufrechterhalten werden. Es ist also für die Modellierung zu klären, ob sich aufgrund der Konstanthaltung der experimentellen Parameter längere Zeit vor dem Puls ein solcher Zustand einstellen konnte. Ein Modell sollte diesen dann auch richtig wiedergeben.

6. Die Ergebnisse der Datenanalyse, s. Abschnitte 4.2 bzw. 4.7, weisen auf eine relativ geringe Ähnlichkeit und relativ hohe Meßunsicherheiten einiger Metabolite hin, die vorwiegend aus dem Zentralstoffwechsel stammen:

Dies kann das Auffinden von Modellen für diesen Stoffwechselabschnitt beeinträchtigen.

Aufgrund der Fragestellung für die Modellierung und der in den Punkten 1-6 aufgeführten Gründe wurde entschieden, hauptsächlich die Intermediate des AAA explizit zu modellieren und die Metabolite des Zentralstoffwechsels nicht in die Modelle zu integrieren. Um trotz dieser starken Vereinfachung sinnvolle Modelle entwickeln zu können, wurden weitere Annahmen für die Eintrittsreaktion in den AAA benötigt. Dafür wurden zwei verschiedene Ansätze gewählt, s. Abschnitt 5.2.2 und 5.2.3.

Zur Entscheidung, ob ein Fließgleichgewichtszustand der intrazellulären Metabolite vor dem Substrat-Puls vorliegt, s. Punkt 5, wurde der Verlauf der Sauerstoffkonzentration im Bioreaktor während der Fermentation in Experiment MU1 untersucht, s. Abb. 3–1^{5.2}. Hier zeigt sich, daß der Wert des Sauerstoffpartialdruckes (pO_2 -Wert) nicht konstant war, sondern in den letzten 22 Minuten vor dem Glucose-Puls einen nahezu linearen Trend von 97% auf 82% aufwies. Da der pO_2 -Wert ein sehr sensibler Parameter für die Stoffwechselaktivität der Zellen ist und in bisherigen Puls-Experimenten, s. (Hirsch 2001; Buchholz 2001), mit einer Fehlertoleranz von $\Delta pO_2 \approx \pm 5\%$ konstant gehalten werden konnte, deutet dies darauf hin, daß kein Fließgleichgewicht vorlag. Daher wurde für die Modelle davon ausgegangen, daß die simulierten Metabolitkonzentrationen vor dem Puls nicht unbedingt konstante Werte aufweisen müssen.

Weiterhin wurde aufgrund von Zusatzinformationen^{5.3} angenommen, daß nur die Metabolite DAH und DHS verstärkt aus der Zelle diffundieren, s. Punkt 1.

^{5.2}Für Experiment MU2 lagen diese Ergebnisse nicht vor.

^{5.3}Ergebnisse aus Vorexperimenten mit dem Stamm *E.coli* 4pF20, s. (Oldiges 2004).

Aus den unter Punkt 4 genannten Gründen erfolgte schließlich die Parameteranpassung des Modells mit über die FFT-Methode gefilterten Trajektorien, wobei ein Krümmungsfaktor von 5 Punkten für die parabolische Sperrfunktion gewählt wurde, s. Abschnitt 4.5.

5.2.2 Modelle mit einem vorgegebenen Flußprofil als Randbedingung

Vereinfachende Annahmen Da die Precursor^{5,4} des AAA nicht explizit modelliert werden sollten, wurden Annahmen benötigt, um einen sinnvollen Wert für den Stofffluß in den AAA vor und nach dem Substrat-Puls abschätzen zu können.

Dazu wurde zunächst die Glucose-Aufnahmerate aus der pO₂-Kurve abgeschätzt, so daß der Fluß der Glucose in die Zelle als Modellparameter festgelegt werden konnte. Infolge des Glucose-limitierten Zustandes vor dem Puls wurde angenommen, daß die im Zufluß zugeführte Glucose-Menge vollständig von den Zellen verbraucht wird und die Biomassekonzentration konstant bleibt. Es gilt also folgende Gleichung:

$$q_{s, \text{cap}, \text{lim.}} = \frac{\text{Gluc}_{in}(t) \cdot F_{in}(t)}{V_R \cdot BTM} \quad (5-1)$$

wobei $q_{s, \text{cap}, \text{lim.}}$ die Glucose-Aufnahmerate im limitierten Zustand und Gluc_{in} die mit der Zuflußrate F_{in} in den Reaktor eingeleitete Glucose-Masse ($\text{Gluc}_{in}(t) \cdot F_{in}(t) = \text{konst.} = 50 \text{ g h}^{-1}$) ist. V_R bezeichnet das Reaktortvolumen (7,66 l) und BTM die Biomassenkonzentration (16,28 gl^{-1}). Die Glucose-Aufnahmerate im Zustand vor dem Puls wird damit auf 0,222 mM / (l s) abgeschätzt.

Nach dem Glucose-Puls kann Gl. (5-1) nicht mehr verwendet werden, da sich dann ein Glucose-Überschuß im Fermenter befindet. Statt dessen wurde die in (Lin et al. 2001) beschriebene Formel zur Abschätzung der maximalen Glucose-Aufnahmerate infolge von Glucose-Pulsen auf Fed-Batch Kultivierungen herangezogen:

$$\left. \begin{aligned} q_{s, \text{cap.}} &= \frac{\text{Gluc} + \int_{t_P}^{t_R} [F_{in}(t) \cdot \text{Gluc}_{in}(t)] dt}{V_R \cdot BTM \cdot (t_R - t_P)}, \\ \Leftrightarrow q_{s, \text{cap.}} &= \frac{\text{Gluc} + (F_{in} \cdot \text{Gluc}_{in}) \cdot (t_R - t_P)}{V_R \cdot BTM \cdot (t_R - t_P)}, \end{aligned} \right\} \text{wenn } F_{in}(t) \cdot \text{Gluc}_{in}(t) = \text{konst.} \quad (5-2)$$

wobei Gluc die Glucosekonzentration des Stoßimpulses und t_R den Zeitpunkt des Abfalls bzw. t_P den Zeitpunkt des erneuten Anstiegs des Sauerstoffsignals ($t_R - t_P = 0,259 \text{ h}$) bezeichnet^{5,5}. Mit Gleichung (5-2) wurde die Glucose-Aufnahmerate nach dem Puls auf einen konstanten Wert von 0,777 mM/(l s) abgeschätzt. Das abgeleitete Glucose-Profil befindet sich in guter Übereinstimmung mit dem Sprungprofil, das in (Chassagnole et al. 2002) über ein DGL-Modell zur Beschreibung eines Substrat-Puls-Experimentes mit *E.coli* berechnet wurde^{5,6}.

Der Fluß der Glucose in die Zelle wurde dann, entsprechend dem gängigen biochemischen Wissen, in einen Fluß in die Glykolyse und den PPP aufgespalten, wobei das Verhältnis der beiden Flüsse als Parameter bei der Anpassung des Modells an die Daten optimiert wurde. Aufgrund von Einschränkungen in der Modellierungssoftware mußte für die Splitting des Flusses eine zusätzliche DGL für einen real

^{5,4}Precursor sind Vorläufermetabolite.

^{5,5}Diese Zeitspanne ist deutlich größer als die Ansprechzeit der Sauerstoffelektrode von ca. 60-80 s.

^{5,6}Das im Modell von (Chassagnole et al. 2002) beobachtete kurzfristige Überspringen der Glucose-Aufnahmerate unmittelbar nach dem Puls konnte mit der hier durchgeführten, einfachen Abschätzung nicht wiedergegeben werden.

nicht existierenden Metaboliten, D, aufgestellt werden:

$$\frac{dD}{dt} = v_{pts} - k_{Glyk} \cdot D - k_{aroF} \cdot D \cdot PEP, \quad (5-3)$$

wobei die Parameter k_{Glyk} und k_{aroF} das Verhältnis der Flüsse zueinander bestimmen. Als zusätzliche Bedingung für die Optimierung galt dann, daß die Konzentration von D gering und konstant sein soll, um das Bodenstein'sche Quasistationaritätsprinzip zu erfüllen (Bodenstein 1913).

Modelle und Ergebnisse Ausgehend von den beschriebenen Annahmen wurde das folgende DGL-System entwickelt, um die Dynamiken der erfaßten Metabolite des Shikimat-Weges zu beschreiben, s. auch Anhang A.4 (S. 152) für eine Übersicht über den Stoffwechselabschnitt:

$$\begin{aligned} \frac{dPEP}{dt} &= -v_{pts} + 2 \cdot v_{d \rightarrow pep} - v_{aroF} - v_{aroA} - v_{pep \rightarrow out} \\ \frac{dDAH}{dt} &= v_{aroF} - v_{aroB} - v_{dah \rightarrow out} \\ \frac{dDHQ}{dt} &= v_{aroB} - v_{aroD} \\ \frac{dDHS}{dt} &= v_{aroD} - v_{aroE,hin} + v_{aroE,rück} - v_{dhs \rightarrow out} \\ \frac{dShik}{dt} &= v_{aroE,hin} - v_{aroE,rück} - v_{aroL,hin} + v_{aroL,rück} \\ \frac{dS3P}{dt} &= v_{aroL,hin} - v_{aroL,rück} - v_{aroA} - v_{s3p \rightarrow out}. \end{aligned} \quad (5-4)$$

In diesem DGL-System wurden nur die hauptsächlich an der Katalyse der entsprechenden Reaktionsschritte beteiligten Isoenzyme, DAHP-Synthase(Tyr) (AroF) und Shikimat-Kinase II (AroL) berücksichtigt. Weiterhin wurde angenommen, daß die Dephosphorylierung von DAHP zu DAH schnell verläuft und sich damit die Metabolite im Gleichgewicht zueinander befinden. Daher wurden die für DAHP bekannten Umsetzungsreaktionen im Modell auf das quantifizierte Intermediat DAH übertragen. Lediglich einfache Kinetiken erster oder zweiter Ordnung wurden den Geschwindigkeitstermen in Gleichung (5-4) zugeordnet, für die keine nennenswerten Regulationsmechanismen bekannt sind. Über die durch die DHQ-Synthase katalysierte Reaktion ist bekannt, daß eine starke Substratinhibierung durch DAH(P) vorliegt (Maitra und Sprinson 1978). Diese Inhibition wurde im Modell durch das Einsetzen einer einfachen Substratinhibitionskinetik berücksichtigt.

Das Modell wurde an die geglätteten Trajektorien aus Experiment MU1 angepaßt, wobei die Ergebnisse der Optimierung, d.h. die simulierten Trajektorien in Abb. 5-1 enthalten sind. Es zeigt sich, daß dieses sehr einfache Modell bereits signifikante Charakteristika der experimentellen Daten wiedergibt. Die Feedback-Inhibierung durch DAH genügt, den 200-fachen Konzentrationsunterschied zwischen DAH und DHQ zu erklären. Die Dynamiken von Shikimat und S3P werden durch das Modell relativ gut nachgebildet. Von einer detaillierten Modellierung einer unmittelbaren Reaktion von DHS und Shikimat, die sich experimentell durch einen kleinen Peak ca. 1 s nach dem Puls abzeichnet, wurde abgesehen. Denn die große DAH-Konzentration, die sich infolge der Feedback-Inhibierung aufbaut und nach dem Puls relativ konstant bleibt, puffert im Modell sofortige Reaktionen der folgenden Metabolite auf den Puls ab. Insofern hätten, um doch noch eine unmittelbare Reaktion von DHQ, Shik und DHS auf den Puls zu erzeugen, weitere Regulationsmechanismen innerhalb des AAA angenommen werden müssen, die

Tab. 5–1 Für die Modelle 1 und 2 in das DGL-System aus Gleichung (5–4) eingesetzte Reaktionskinetiken zur Beschreibung der Datensätze aus den Experimenten MU1 und MU2; s. Anhang G.1 für die Parameterwerte

Fluß	Kinetik in Modell 1	Kinetik in Modell 2
$v_{pep \rightarrow out}$	$k_{pep \rightarrow out} \cdot PEP$	$k_{pep \rightarrow out} \cdot PEP$
v_{aroF}	$k_{aroF} \cdot D \cdot PEP$	$k_{aroF} \cdot D \cdot PEP$
v_{aroA}	$k_{aroA} \cdot S3P \cdot PEP$	$\frac{V_{max,aroA} \cdot PEP \cdot S3P}{(K_{m,aroA,S3P} + S3P) \cdot (K_{m,aroA,PEP} + PEP)}$
v_{aroB}	$\frac{V_{max,aroB} \cdot DAH}{\left(K_{m,aroB} + DAH \cdot \left(1 + \frac{DAH}{K_{i,aroB}}\right)\right)}$	$\frac{V_{max,aroB} \cdot DAH}{\left(K_{m,aroB} + DAH \cdot \left(1 + \frac{DAH}{K_{i,aroB}}\right)\right)}$
$v_{dah \rightarrow out}$	$k_{dah \rightarrow out} \cdot DAH$	$k_{dah \rightarrow out} \cdot DAH$
v_{aroD}	$k_{aroD} \cdot DHQ$	$\frac{V_{max,aroD} \cdot DHQ}{(K_{m,aroD} + DHQ)}$
$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out} \cdot DHS$	$k_{dhs \rightarrow out} \cdot DHS$
$v_{aroE,hin}$	$k_{aroE} \cdot DHS$	$\frac{V_{max,aroE} \cdot DHS \cdot (K_{eq,aroE} - NADP)}{(K_{m,aroE,NADP} + (K_{eq,aroE} - NADP)) \cdot (K_{m,aroE,DHS} + DHS)}$
$v_{aroE,rück}$	–	$k_{aroE,rück} \cdot Shik \cdot NADP$
$v_{aroL,hin}$	$k_{aroL,hin} \cdot Shik \cdot ATP$	$\frac{V_{max,aroL} \cdot Shik \cdot ATP}{(K_{m,aroL,Shik} + Shik) \cdot (K_{m,aroL,ATP} + ATP)}$
$v_{aroL,rück}$	–	$k_{aroL,rück} \cdot S3P \cdot ADP$
$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out} \cdot S3P$	$k_{s3p \rightarrow out} \cdot S3P$

bisher noch nicht bekannt sind. Nur der Metabolit PEP konnte mit dem einfachen Modell nicht adäquat beschrieben werden. Hier ist zu vermuten, daß evtl. die anaplerotische Reaktion von Oxalacetat zu PEP oder eine Verlagerung des PEP / Pyr-Gleichgewichts der Pyruvat-Kinase-Reaktion im Experiment die PEP-Konzentration aufrecht erhält, während im Modell aufgrund des Fehlens entsprechender experimenteller Informationen eine Neubildung von PEP nur aus dem Fluß durch die Glykolyse resultiert.

Bei Betrachtung der Flußverteilung, s. Abb. 5–1 (rechts) auf S. 74, zeigt sich für die DAH-Konvertierung ein stark geändertes Flußprofil und ein geringerer Fluß in Bezug auf die Eintrittsreaktion in den AAA. Damit limitiert die DHQ-Synthase offenkundig den Stofffluß durch den AAA. Daher wurde in Experiment MU2 ein Stamm verwendet, in dem das Enzym überexprimiert wurde. Dies wurde simulativ dadurch nachgebildet, daß das gleiche Modell (Modell 1) diesmal an die Daten aus Experiment MU2 angepaßt wurde, wobei nur der entsprechende V_{max} -Wert der DHQ-Synthase um den Faktor 10 erhöht und die übrigen Parameterwerte übernommen wurden. Die Ergebnisse der Simulation sind in Abbildung 5–2 auf S. 74 dargestellt.

Die simulierten Trajektorien von DHQ, DHS und Shikimat stimmen tendenziell relativ gut mit den experimentellen Daten überein. Damit ist Modell 1 bereits in der Lage, den Einfluß der Überexpression der DHQ-Synthase auf den AAA adäquat wiederzugeben. Die schlechtere Anpassung von S3P zeigt jedoch, daß nach Entfernung des in Experiment MU1 wesentlichen Regulationsmechanismus evtl. der Einfluß weiterer Regulationsfaktoren zu Tage tritt, deren Identifikation noch aussteht. Die intrazelluläre Konzentration von DAH war in diesem Experiment zu gering, als daß sie quantitativ hätte angegeben werden können. Daher wurde die simulierte DAH-Trajektorie nicht an die Daten angepaßt. Es galt lediglich als zusätzliche Einschränkung bei der Optimierung, daß simulierte DAH-Konzentration einen Grenzwert von 10 μ M nicht überschreiten sollte. Der Metabolit PEP konnte wiederum nicht sonderlich gut durch das Modell beschrieben werden. Es wurden offenbar doch bzgl. der Glucose-Aufnahme, der

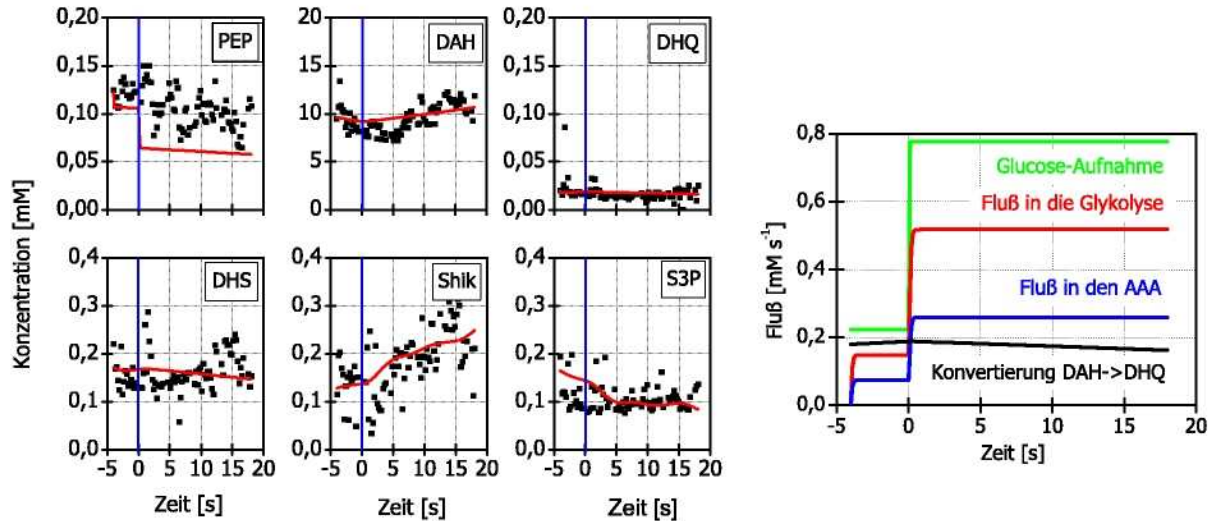


Abb. 5-1 Simulationsergebnisse des Modells 1 zur Beschreibung der Trajektorien aus Experiment MU1

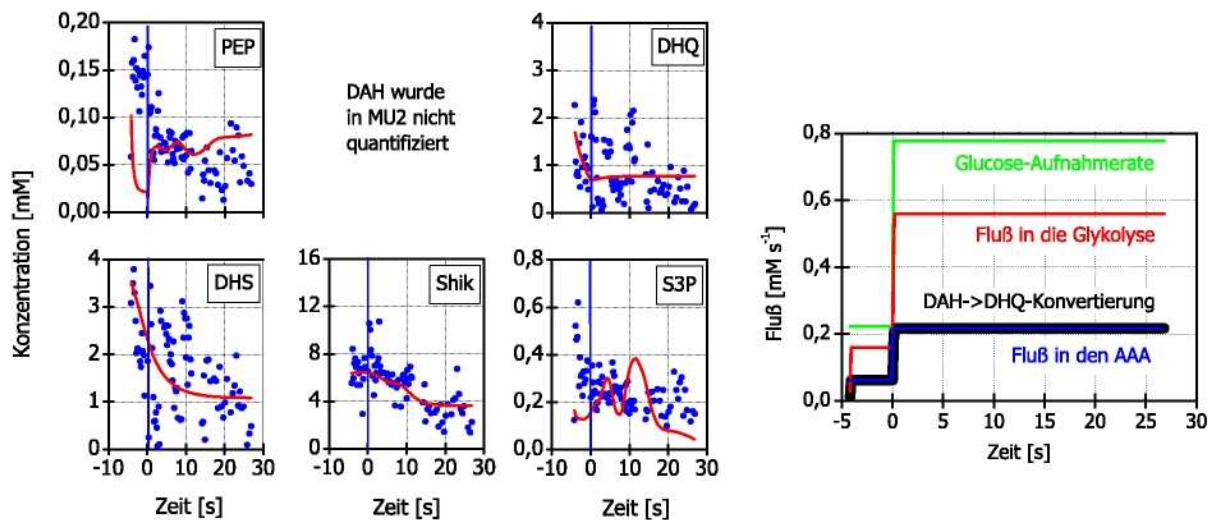


Abb. 5-2 Simulationsergebnisse des Modells 1 zur Beschreibung der Trajektorien aus Experiment MU2

Glykolyse und des Citronensäurezyklus so stark vereinfachende Annahmen gemacht, daß es mit dem entwickelten Modell nicht möglich ist, eine bessere Anpassungsgüte zu erzielen.

Im Vergleich zu den Simulationsergebnissen für Experiment MU1 bleiben die Flüsse in die Glykolyse bzw. den AAA in ihrer Größe erhalten, s. Abb. 5-2 (rechts). Der Fluß über die DHQ-Synthese entspricht jedoch jetzt dem Fluß in den AAA und das Flußprofil ändert sich nicht. Daran zeigt sich, daß die durch AroB verursachte Limitierung des Stoffflusses durch den AAA nicht mehr existiert.

Insgesamt ist jedoch die Anpassungsqualität von Modell 1 in Bezug auf einige Metabolite aus dem Puls-Experiment MU2 noch nicht befriedigend. Gründe hierfür könnten sein, daß (a) die Formulierung des Modellteils für den AAA noch nicht detailliert genug ist, oder daß (b) die Modelle im Bereich des Zentralstoffwechsels zu stark vereinfacht wurden.

Um den ersten Punkt genauer zu untersuchen, wurden für eine realistischere Beschreibung des AAA in Modell 2 alle Kofaktoren der Reaktionen berücksichtigt, teilweise reversible Reaktionsschritte eingesetzt und statt der Geschwindigkeitsgesetze erster bzw. zweiter Ordnung einfache enzymkinetische

Gleichungen verwendet, s. Tab. 5–1 (rechte Spalte). Auch der Einfluß von NADPH, das als Kofaktor für die Reduktion von DHS benötigt wird, wurde im Modell berücksichtigt, obwohl es experimentell nicht quantifiziert werden konnte. Dazu wurde ebenso wie in (Wolf und Heinrich 2000; Visser 2001) angenommen, daß die Summe der NADP- und NADPH-Konzentration über die Zeit hinweg konstant ist: $\text{NADP} + \text{NADPH} = \text{konst.}$, $\text{NADH/NAD} = K_{eq,aroE}$.

Die Verfeinerung des Modells ergab jedoch keine Verbesserung bzgl. der Anpassungsgüte, d.h. die simulierten Trajektorien haben den gleichen Verlauf wie in Abb. 5–2. Dies ist möglicherweise darauf zurückzuführen, daß die Dynamiken im AAA in größerem Ausmaß durch die Eingangssubstrate PEP und E4P oder durch Effektoren aus dem Zentralstoffwechsel verursacht werden.

5.2.3 Modelle mit der Vorgabe von Konzentrationsverläufen als Randbedingung

Da es trotz der Einbeziehung wesentlicher Regulationsmechanismen innerhalb des Aromatenbiosyntheseweges mit Modell 2 nicht gelungen ist, die Metabolittrajektorien besser, als in Abb. 5–2 dargestellt, anzupassen, wurden die starken Vereinfachungen des Modells im Bereich des Zentralstoffwechsels als wesentliche Ursache für eine noch nicht zufriedenstellende Simulation angesehen. Aus den in Abschnitt 5.2.1 genannten Gründen ist es jedoch kaum möglich, ein detailliertes Modell für die Glykolyse und den PPP aufzustellen.

Tab. 5–2 Für die Modelle 4 und 5 in das DGL-System aus Gleichung (5–6) eingesetzte Reaktionskinetiken; s. Anhang G.1 für die Parameterwerte

Fluß	Kinetik in Modell 4	Kinetik in Modell 5
$v_{c5p \rightarrow dhq}$	$k_1 \cdot \text{C5P}_{\text{exp}}(t) \cdot \text{PEP}_{\text{exp}}(t) - k_2 \cdot \text{DHQ}$	$k_1 \cdot \text{C5P}_{\text{exp}}(t) \cdot \text{PEP}_{\text{exp}}(t) - k_2 \cdot \text{DHQ}$
v_{aroD}	$k_{aroD} \cdot \text{DHQ}$	$k_{aroD} \cdot \text{DHQ}$
$v_{dhq \rightarrow out}$	$k_{dhq \rightarrow out} \cdot \text{DHQ}$	$k_{dhq \rightarrow out} \cdot \text{DHQ}$
$v_{aroE, hin}$	$\frac{V_{max,aroE} \cdot \text{DHS} \cdot (K_{eq,aroE} - \text{NADP})}{(K_{m,aroE,NADP} + (K_{eq,aroE} - \text{NADP})) \cdot (K_{m,aroE,DHS} + \text{DHS})}$	” ”
$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out} \cdot \text{DHS}$	$k_{dhs \rightarrow out} \cdot \text{DHS}$
$v_{aroL, hin}$	$k_{aroL, hin} \cdot \text{Shik} \cdot \text{ATP}$	$\frac{V_{max,aroL} \cdot \text{Shik}}{(K_{m,aroL,Shik} + \text{Shik})}$
$v_{aroL, rück}$	$k_{aroL, rück} \cdot \text{S3P} \cdot \text{ADP}$	–
v_{aroA}	$k_{aroA} \cdot \text{S3P} \cdot \text{PEP}$	$k_{aroA} \cdot \text{S3P} \cdot \text{PEP}$
$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out} \cdot \text{S3P}$	$k_{s3p \rightarrow out} \cdot \text{S3P}$

Daher wurden die experimentell bestimmten Konzentrationen der an der Eintrittsreaktion in den AAA beteiligten Edukte, d.h. PEP und die Summenkonzentration der Zucker-5-Phosphate, stellvertretend für den nicht bestimmten Metaboliten E4P, als zeitabhängige Modellparameter eingesetzt. Ziel war zu untersuchen, wie sich die durch diese Metabolite eingetragenen Dynamiken im AAA fortpflanzen. Dazu wurde zunächst Modell 1 nur dahingehend abgewandelt, daß D, PEP und DAH nicht mehr als abhängige Variablen in das DGL-System integriert wurden. Damit resultierte ein drittes Modell, indem die DGL für DHQ aus Modell 1, s. Gleichung (5–4) und Tab. 5–1, wie folgt umformuliert wurde:

$$\frac{d \text{DHQ}}{dt} = k_1 \cdot \tilde{\text{C5P}} \cdot \tilde{\text{PEP}} - k_5 \cdot \text{DHQ}. \quad (5-5)$$

Ebenso wurde keine DGL mehr für DAH formuliert, da der Metabolit in Experiment MU2 nicht quantifiziert werden konnte. Es wurde statt dessen angenommen, daß sich die Dynamiken der Edukte des AAA direkt in der Konzentration von DHQ widerspiegeln. Eine verbesserte Anpassungsqualität des

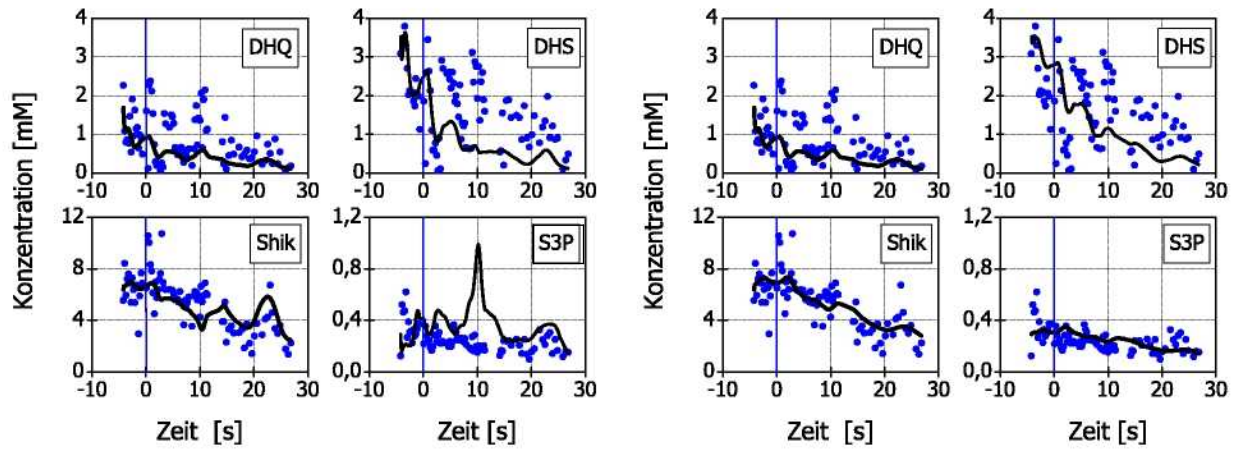


Abb. 5–3 Simulationsergebnisse der Modelle 4 (linke Seite) und Modell 5 (rechte Seite) zur Beschreibung der Trajektorien aus Experiment MU2

Modells 3 ließ sich jedoch noch nicht feststellen.

Erst nach Berücksichtigung aller Kofaktoren der AAA-Reaktionen und bei Anwendung komplexerer Enzymkinetiken konnten in einem vierten Modell, s. die folgende Gleichung und Tab. 5–2 (linke Spalte), die experimentellen Trajektorien relativ gut angepaßt werden, s. Abb. 5–3.

$$\begin{aligned}
 \frac{d \text{ DHQ}}{dt} &= v_{c5p \rightarrow dhq} - v_{aroD} - v_{dhq \rightarrow out} \\
 \frac{d \text{ DHS}}{dt} &= v_{aroD} - v_{aroE} - v_{dhs \rightarrow out} \\
 \frac{d \text{ Shik}}{dt} &= v_{aroE} - v_{aroL, hin} + v_{aroL, rück} \\
 \frac{d \text{ S3P}}{dt} &= v_{aroKL, hin} - v_{aroKL, rück} - v_{aroA} - v_{s3p \rightarrow out}.
 \end{aligned}
 \tag{5-6}$$

Wenn die experimentell bestimmten Trajektorien von Metaboliten aus den Daten eingelesen werden, anstatt sie durch DGL zu simulieren, ist zu beachten, daß es möglicherweise zu unrealistisch hohen Abflüssen aus diesen zeitabhängigen Parametern kommen kann. Um die Gültigkeit einer solchen Simulation zu prüfen, sollte daher nach der Optimierung in jedem Fall eine Kontrolle der Flußverteilung erfolgen. Im Fall von Modell 4 zeigte sich ein unphysiologisch hoher Wert von ca. 25 mM / s für den Fluß in den AAA. Weiterhin erschien das Modell 4 hinsichtlich des Peaks in der simulierten S3P-Trajektorie bei ca. 10 s noch verbesserungswürdig.

Daher wurde in einem fünften Modell eine Limitierung des Flusses in den AAA festgelegt, d.h. der Fluß sollte eine optimierte obere Grenze von 4 mM / s nicht überschreiten, was in etwa dem Vierfachen der berechneten Glucose-Aufnahmerate entspricht, s. Gleichung (5–2). Als Ursache des Peaks in der S3P-Trajektorie wurde die ATP-Trajektorie identifiziert, die zu dieser Zeit einen ähnlichen Verlauf aufweist, s. Anhang C.3. Daraus läßt sich schließen, daß offenbar die Shikimat-Kinase nicht in dem Ausmaß durch ATP limitiert ist, wie in Modell 4 angenommen wurde. Aus diesem Grund wurde in dem verbesserten fünften Modell bewußt der Einfluß dieses Kofaktors vernachlässigt, s. Tab. 5–2 (rechte Spalte).

Die Ergebnisse der Simulation sind in Abb. 5–3 dargestellt. Eine recht hohe Anpassungsgüte ist für

alle vier modellierten Metabolite erkennbar. Somit ist zu vermuten, daß vorwiegend die im Zentralstoffwechsel beobachteten Dynamiken sich auf den Verlauf der AAA-Intermediate auswirken. Es ist bei der Modellierung der Experimente daher nicht nötig, außer der Substratinhibierung der DHQ-Synthase in MU1 noch weitere Regulationsmechanismen von Metaboliten des AAA vorauszusetzen.

5.2.4 Zusammenfassung der Modellfamilie

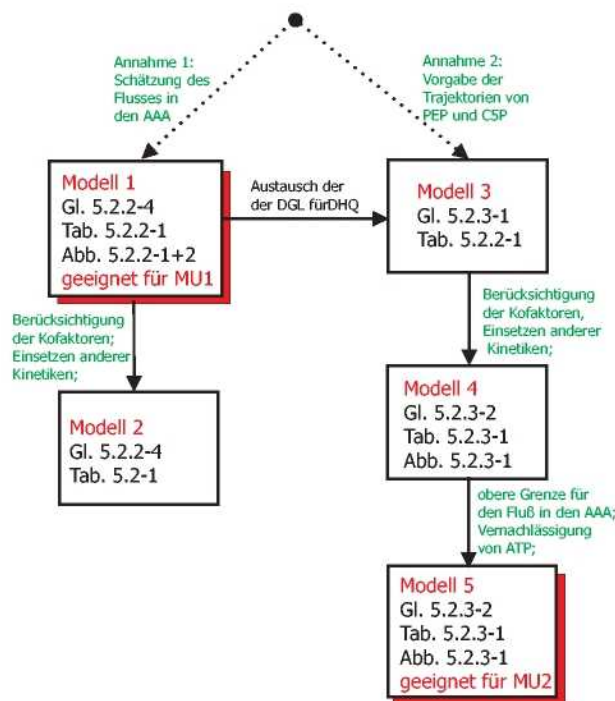


Abb. 5–4 Modellfamilie zur Beschreibung der Aromatenbiosynthese in *E.coli*-Mutanten

Ziel der Untersuchungen war es, die experimentell bestimmten Konzentrationsniveaus und dynamischen Verläufe der AAA-Intermediate durch DGL-Modelle wiederzugeben. Zu diesem Zweck wurde unter der Voraussetzung von verschiedenen vereinfachenden Annahmen eine Familie aus insgesamt fünf Modellen erstellt, die gegenseitig aufeinander aufbauen. Ausgehend von dem Ursprungsmodell, Modell 1, wurden jeweils nur bestimmte Bausteine eines Modells ausgetauscht, hinzugenommen oder weggelassen, um das nächste Modell zu erzeugen, s. Abb. 5–4.

Anstatt an dieser Stelle jedoch noch detailliertere Stoffwechselmodelle für die Experimente MU1 und MU2 zu erstellen, sollten zunächst die Ursachen der Meßdatenstreuungen geklärt werden. Daher wird eine Reproduktion der Experimente vorgeschlagen, um eindeutige Trends in den gemessenen Trajektorien zu identifizieren, die im Anschluß durch ein Modell wiedergegeben werden können.

5.3 Modelle für die Glykolyse und den PPP von *E.coli* K12

5.3.1 Vorstellung eines Modells

Die Modelle zur Beschreibung der Daten aus Experiment WT1 sind analog zu der Vorgehensweise, wie sie in Abschnitt 5.2 geschildert wurde, entstanden. D.h. es existieren verschiedene Modellfamilien, s. Anhang G, die jeweils von einem einfachen Ursprungsmodell ausgehen und meist mehrere Modelle

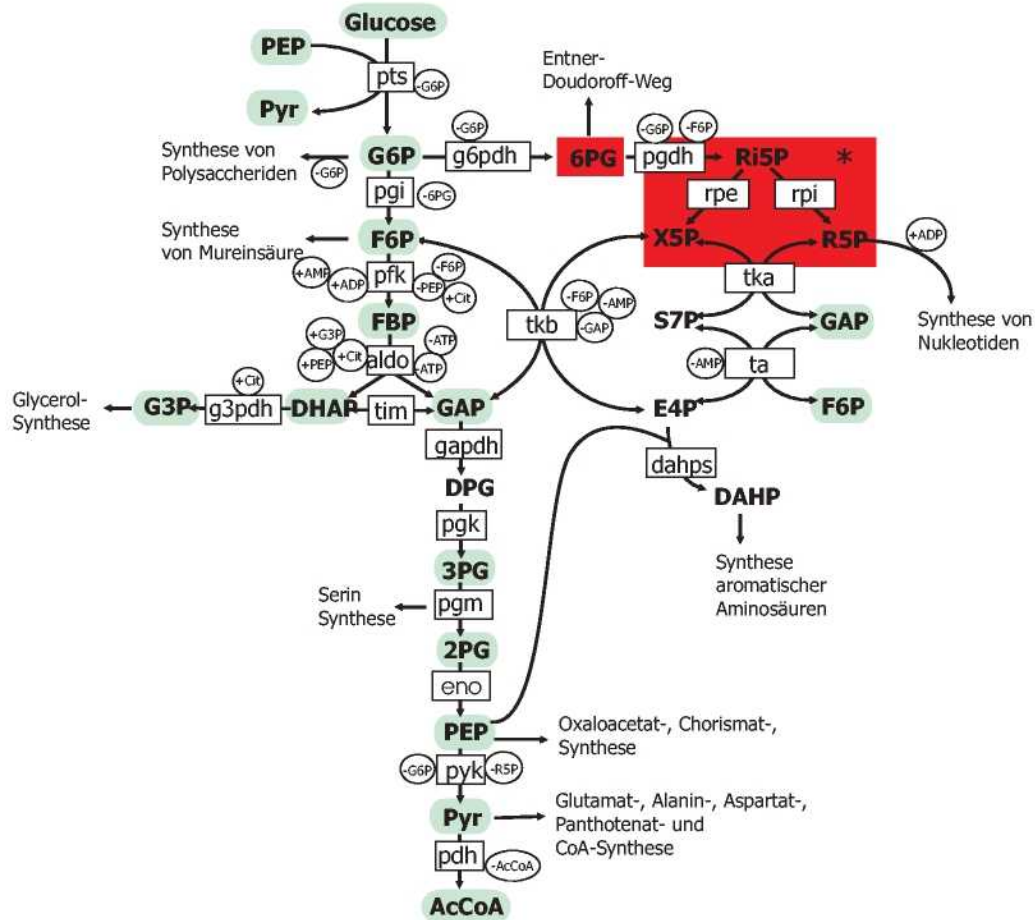


Abb. 5–5 Stoffwechselnetzwerk der Glykolyse und des PPP in einem *E.coli*-Wildstamm; abgerundete Kästen: über enzymatische Assays, eckige Kästen: über die ESI-LC-MS-Technik quantifizierte Metabolite; eckiger Kasten mit Stern: die Summenkonzentration der Intermediate wurde erfaßt; umrandete Kästen: Enzyme, umrandete Ellipsen: bekannte Effektoren der Reaktionsschritte

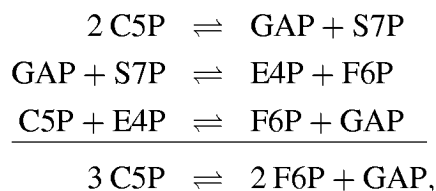
mit einer relativ guten Anpassungsgüte hervorbringen. Eines dieser Modelle soll im Folgenden näher beschrieben werden, wobei die zur Modellierung benötigten, wesentlichen Vorinformationen, d.h. das zugrundeliegende Reaktionsnetzwerk und die bekannten Effektoren der Enzymreaktionen zusammen mit einer Angabe der vorhandenen Meßdaten in Abb. 5–5 zusammengefaßt werden.

Annahmen zur Generierung des DGL-Systems Im Zuge der Modellerstellung war zunächst zu entscheiden, für welche Metabolite aus Abb. 5–5 im Modell Massenbilanzen aufgestellt werden sollten. Dazu wurde beschlossen, nur experimentell bestimmte Intermediate über DGL zu bilanzieren, da die Modelle anhand der Meßdaten auch praktisch identifizierbar sein sollten, um eindeutige Aussagen aus ihnen ableiten zu können, s. Abschnitt 3.4.2. Zusätzlich wurde auf eine Modellierung der Nukleotidkonzentrationen verzichtet. Diese sind an sehr vielen Stoffwechselreaktionen beteiligt, so daß ihre zeitliche Konzentrationsänderung vermutlich nicht adäquat durch ein Modell nur für die Glykolyse und den PPP beschrieben werden kann. Statt dessen wurde ihr Konzentrationsverlauf als zeitabhängiger Modellparameter aus den Meßdaten eingelesen und, wenn nötig, eine konstante Summenkonzentration der oxidierten und reduzierten bzw. der verschiedenen Phosphorylierungs-Formen angenommen. Auch von den verbleibenden Trajektorien wurden einige, nämlich die der Metabolite 3PG, 2PG, AcCoA und OxAc nicht durch das Modell nachgebildet, weil bei ihnen die Anzahl der erfaßten Meßpunkte

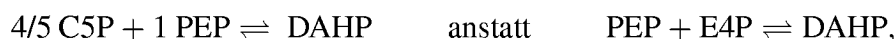
oder die Datenqualität zu gering erachtet wurde, s. Anhang C.1. Dies erforderte Vereinfachungen der stöchiometrischen Gleichungen an den betreffenden Stellen des Stoffwechselnetzwerkes.

Infolgedessen wurde z.B. die lineare Reaktionssequenz $\text{GAP} \rightleftharpoons \text{DPG} \rightleftharpoons \text{3PG} \rightleftharpoons \text{2PG} \rightleftharpoons \text{PEP}$, in der Glykolyse übereinstimmend mit (Heinrich und Rapoport 1974b; Richter et al. 1975; Theobald et al. 1997a) durch eine Reaktion: $\text{GAP} \rightarrow \text{PEP}$ ersetzt, da sich die Intermediate DGP, 3PG, 2PG bekanntermaßen rasch ineinander umwandeln und sich damit vor und nach dem Substrat-Puls im Fließgleichgewicht befinden (Spring und Wold 1971; Molnar und Vas 1993; Schmidt et al. 1995). Ferner werden sie nicht in größerer Menge als Precursor für die Biosynthese von hochmolekularen Verbindungen benötigt (Kanehisa und Goto 2000; Karp et al. 2000).

Innerhalb des PPP konnte experimentell nur eine Summenkonzentration der Zucker-5-Phosphate Ri5P, R5P und X5P erfaßt werden, deren zeitlicher Verlauf durch eine entsprechende Modellgröße C5P simuliert wurde. Dazu galt wiederum in Übereinstimmung mit (Schuster et al. 1988; Vaseghi et al. 1999; Chassagnole et al. 2002) die Annahme, daß diese Intermediate rasch miteinander reagieren, und sich also wahrscheinlich zwischen ihnen ein Fließgleichgewicht einstellt. Um die nicht bestimmten Intermediate S7P und E4P aus dem Modell eliminieren zu können, wurde weiterhin die Transketolase und die Transaldolase-Reaktion folgendermaßen vereinfacht:



wobei die Eintrittsreaktion in den Aromatenbiosyntheseweg von E4P auf die Modellgröße C5P übertragen wurde:



da 4 C5P-Moleküle 5 E4P-Molekülen in Bezug auf die Anzahl von C-Atomen entsprechen, s. auch Tab. E-1 (S. 173) zu den Vereinfachungen. Ansonsten wurde die bekannte Stöchiometrie der Reaktionen

Tab. 5-3 DGL-System eines Modells zur Beschreibung des Zentralstoffwechsels in *E.coli* K12

$\frac{d \text{ G6P}}{dt}$	$= v_{pts} - v_{pgi} - v_{glp} - v_{g6pdh}$
$\frac{d \text{ F6P}}{dt}$	$= v_{pgi} - v_{pfk} - v_{mur} + v_{fbpase} + 2 \cdot v_{tkata}$
$\frac{d \text{ FBP}}{dt}$	$= v_{pfk} - v_{aldo} - v_{fbp \rightarrow bm} - v_{fbpase}$
$\frac{d \text{ GAP}}{dt}$	$= v_{aldo} - v_{gap \rightarrow pep} - v_{ser} + v_{tim} + v_{tkata}$
$\frac{d \text{ DHAP}}{dt}$	$= v_{aldo} - v_{g3pdh} - v_{tim}$
$\frac{d \text{ G3P}}{dt}$	$= v_{g3pdh} - v_{gly}$
$\frac{d \text{ PEP}}{dt}$	$= v_{gap \rightarrow pep} - v_{pyk} - v_{pts} - v_{dahps} - v_{pep \rightarrow bm} - 5 \cdot v_{dahps}$
$\frac{d \text{ Pyr}}{dt}$	$= v_{pyk} - v_{pdh} + v_{pts}$
$\frac{d \text{ 6PG}}{dt}$	$= v_{g6pdh} - v_{pgdh} - v_{pghl}$
$\frac{d \text{ C5P}}{dt}$	$= v_{pgdh} - 3 \cdot v_{tkata} - v_{rppk} - 4 \cdot v_{dahps}$

auch bzgl. etwaiger Kofaktoren beibehalten und die von den Intermediaten F6P, FBP, GAP, PEP, G3P

und C5P ausgehenden Abzweigungen in die verschiedenen Biosynthesewege, s. Abb. 5–5 (Pramanik und Keasling 1997), im Modell durch jeweils einen Abfluß pro Metabolit berücksichtigt. Damit ergibt sich das in Tab. 5–3 dargestellte DGL-System.

Annahmen für die Zuordnung von Enzymkinetiken zu den Geschwindigkeitstermen Nach der Erstellung des DGL-Systems wurden kinetische Terme für die einzelnen Reaktionsgeschwindigkeiten aus Tab. 5–3 eingesetzt. Hierzu war für jeden Reaktionsschritt eine passende algebraische Form der betreffenden Enzymkinetik zu bestimmen, s. Tab. 5–4 für die Zusammenstellung verwendeten Kinetiken. Um wegen der praktischen Identifizierbarkeit der Modelle die Zahl der Modellparameter gering zu halten, wurden einfache bekannte Grundformen von kinetischen Gleichungen gewählt und der Einfluß von Effektoren über zusätzliche multiplikative Terme berücksichtigt, insbesondere wenn keine weiteren Informationen über die Art der Inhibition, s. Anhang D, vorhanden waren.

Tab. 5–4 Reaktionskinetische Terme für das Modell aus Tab. 5–3

Phosphotransferase System

$$v_{pts} = k_{f,pts} \text{Gluc}^{n_{pts,1}} \text{PEP}^{n_{pts,2}} - k_{b,pts} \text{G6P}^{n_{pts,3}} \text{Pyr}^{n_{pts,4}} \quad (5-7)$$

Phosphoglucoisomerase

$$v_{pgi} = k_{f,pgi} \text{G6P}^{n_{pgi,1}} - k_{b,pgi} \text{F6P}^{n_{pgi,2}} \quad (5-8)$$

Synthese von Polysacchariden

$$v_{glp} = \frac{V_{f,glp} \text{G6P} \cdot \text{ATP}}{(K_{m,G6P,glp} + \text{G6P}) \cdot (K_{m,ATP,glp} + \text{ATP})} \cdot \frac{1}{(\text{G6P} + K_{i,G6P,glp})^{n_{glp,1}}} \quad (5-9)$$

Phosphofructokinase

$$v_{pfk} = \frac{2b \cdot D \cdot e \cdot \text{F6P} (1 + \text{FBP})^2 \left(1 + \frac{\text{F6P}}{e+1}\right)}{(e+1) \left(L (1 + \text{F6P}c)^2 + (1 + \text{FBP})^2 \left(1 + \frac{\text{F6P}}{e+1}\right)^2 \right)} \cdot \frac{\text{ATP}}{K_{m,ATP,pfk} + \text{ATP}} \cdot \frac{\text{ADP}^{n_{pfk,1}}}{(\text{ADP}^{n_{pfk,1}} + K_{i,ADP,pfk}^{n_{pfk,1}})} \cdot \frac{\text{F6P}^{n_{pfk,2}}}{(\text{F6P}^{n_{pfk,2}} + K_{i,\text{F6P,pfk}}^{n_{pfk,2}})} \cdot \frac{1}{(K_{i,\text{PEP,pfk}}^{n_{pfk,3}} + \text{PEP}^{n_{pfk,3}})} \quad (5-10)$$

Synthese von Mureinsäure

$$v_{mur} = \frac{V_{f,mur} \text{F6P} \cdot \text{ATP}}{(K_{m,\text{F6P,mur}} + \text{F6P}) (K_{m,\text{ATP,mur}} + \text{ATP})} \quad (5-11)$$

Fructose-1,6-Bisphosphatase

$$v_{fbpase} = \frac{V_{f,fbpase} \text{FBP}^{n_{fbpase}}}{K_{m,fbpase}^{n_{fbpase}} + \text{FBP}^{n_{fbpase}}} \quad (5-12)$$

Aldolase

$$v_{aldo} = \frac{V_{\max,aldo} \text{FBP}^{n_{aldo,1}}}{K_{m,aldo}^{n_{aldo,1}} + \text{FBP}^{n_{aldo,1}}} \cdot \left(\frac{\text{G3P}}{K_{\text{G3P,aldo}}} \right)^{n_{aldo,2}} \quad (5-13)$$

Fluß: FBP → Biomasse

$$v_{fbp \rightarrow bm} = \frac{V_{f,fbp \rightarrow bm} \text{FBP}^{n_{fbp \rightarrow bm}}}{K_{m,fbp \rightarrow bm}^{n_{fbp \rightarrow bm}} + \text{FBP}^{n_{fbp \rightarrow bm}}} \quad (5-14)$$

Triosephosphat-Isomerase

$$v_{tim} = \frac{\left(\frac{V_{f,tim} \text{DHAP}}{K_{m,DHAP,tim}} \right) - \left(\frac{V_{b,tim} \text{GAP}}{K_{m,GAP,tim}} \right)}{1 + \left(\frac{\text{DHAP}}{K_{m,DHAP,tim}} \right) + \left(\frac{\text{GAP}}{K_{m,GAP,tim}} \right)} \quad (5-15)$$

Tab. 5–4 Reaktionskinetische Terme für das Modell aus Tab. 5–3*Glycerol-3-Phosphat-Dehydrogenase*

$$v_{g3pdh} = \frac{V_{f,g3pdh} \text{DHAP}^{n_{g3pdh,1}}}{K_{m,DHAP,g3pdh} + \text{DHAP}^{n_{g3pdh,1}}} \cdot \left(\frac{\text{Citrat}}{K_{a,Cit,g3pdh}} \right)^{n_{g3pdh,2}} \quad (5-16)$$

Glycerol-Synthese

$$v_{gly} = \frac{V_{f,gly} \text{G3P}^{n_{gly,1}}}{K_{m,G3P,gly} + \text{G3P}^{n_{gly,1}}} \cdot \left(\frac{\text{F6P}}{K_{a,F6P,gly}} \right)^{n_{gly,2}} \quad (5-17)$$

Summenreaktion: GAP → PEP

$$v_{gap \rightarrow pep} = \frac{V_{f,gap \rightarrow pep} \text{GAP} \cdot \text{NAD} \cdot \text{ADP}}{(K_{m,GAP,gap \rightarrow pep} + \text{GAP}) \cdot (K_{m,NAD,gap \rightarrow pep} + \text{NAD}) \cdot (K_{m,ADP,gap \rightarrow pep} + \text{ADP})} \quad (5-18)$$

Serin-Synthese

$$v_{ser} = \frac{V_{f,ser} \text{GAP}^{n_{ser}}}{K_{m,ser} + \text{GAP}^{n_{ser}}} \quad (5-19)$$

Pyruvat-Kinase

$$v_{pyk} = \frac{V_{f,pyk} \left(\text{PEP} \cdot \text{ADP} - \frac{\text{Pyr} \cdot \text{ATP}}{K_{eq,pyk}} \right)}{A + B} \cdot \left(\frac{K_{i,G6P,pyk}}{\text{G6P}} \right)^{n_{pyk,1}} \cdot \left(\frac{K_{i,C5P,pyk}}{\text{C5P}} \right)^{n_{pyk,2}} \quad (5-20)$$

$$A = \text{PEP} \cdot \text{ADP} \cdot \left(1 + \frac{\text{Pyr}}{K_{i,Pyr,pyk}} \right) + K_{m,ADP,pyk} (\text{PEP} + K_{i,PEP,pyk}) + K_{m,PEP,pyk} \text{ADP}$$

$$B = \frac{V_f}{V_r K_{eq}} \left(K_{m,ATP,pyk} \text{Pyr} \left(1 + \frac{\text{PEP}}{K_{i,PEP,pyk}} \right) + \text{ATP} \cdot C \right)$$

$$C = \left[K_{m,Pyr,pyk} \left(1 + \frac{K_{m,PEP,pyk} \text{ADP}}{K_{i,PEP,pyk} K_{m,ADP,pyk}} \right) + \text{Pyr} \left(1 + \frac{\text{ADP}}{K_{i,ADP,pyk}} \right) \right]$$

Fluß: PEP → Biomasse

$$v_{pep \rightarrow bm} = \frac{V_{f,pep \rightarrow bm} \text{PEP} \cdot \text{ATP}}{K_{m,ATP,pep \rightarrow bm} \cdot (K_{i,PEP,pep \rightarrow bm} + \text{PEP}) + \text{ATP} \cdot (K_{m,PEP,pep \rightarrow bm} + \text{PEP})} \quad (5-21)$$

Pyruvatdehydrogenase

$$v_{pdh} = \frac{V_{f,pdh} \text{Pyr} \cdot \text{NAD}}{(K_{m,Pyr,pdh} + \text{Pyr}) (K_{m,NAD,pdh} + \text{NAD})} \frac{1}{(K_{i,AcCoA,pdh} + \text{AcCoA})^{n_{pdh}}} \quad (5-22)$$

Glucose-6-Phosphat-Dehydrogenase

$$v_{g6pdh} = \frac{V_{f,g6pdh} \text{G6P} \cdot \text{NADP}}{K_{m,NADP,g6pdh} (\text{G6P} + K_{i,G6P,g6pdh}) \cdot A + \text{NADP} (\text{G6P} + K_{m,G6P,g6pdh})} \quad (5-23)$$

$$A = \left(1 + \frac{K_{eq,g6pdh} - \text{NADP}}{K_{i,NADP,g6pdh}} \right)$$

6-Phosphogluconatdehydrogenase

$$v_{pgdh} = \frac{V_{f,pgdh} \left(6\text{PG} \cdot \text{NADP} - \frac{\text{C5P}}{K_{eq,pgdh}} \right)}{A + B} \cdot \left(\frac{K_{i,G6P,pgdh}}{\text{G6P}} \right)^{n_{pgdh,1}} \quad (5-24)$$

$$A = 6\text{PG} \cdot \text{NADP} + K_{m,6PG,pgdh} \text{NADP} + K_{m,NADP,pgdh} 6\text{PG}$$

$$B = \frac{V_{f,pgdh}}{V_{r,pgdh} K_{eq,pgdh}} \left(K_{m,C5P,pgdh} + \text{C5P} \left(1 + \frac{6\text{PG}}{K_{i,6PG,pgdh}} \right) \right) + \left(\frac{K_{i,F6P,pgdh}}{\text{F6P}} \right)^{n_{pgdh,2}}$$

Tab. 5–4 Reaktionskinetische Terme für das Modell aus Tab. 5–3*Entner–Doudoroff–Weg*

$$v_{pghl} = \frac{V_{f,pghl} 6PG^{n_{pghl,1}}}{K_{m,pghl}^{n_{pghl,1}} + 6PG^{n_{pghl,1}}} \cdot \left(\frac{F6P}{K_{a,F6P,pghl}} \right)^{n_{pghl,2}} \quad (5-25)$$

Transaldolase-Transketolase-Summenreaktion

$$v_{tkata} = \frac{V_{f,tkata} \left(C5P - \frac{GAP \cdot F6P}{K_{eq,tkata}} \right)}{\left(K_{m,C5P,tkata} + C5P \left(1 + \frac{GAP}{K_{i,GAP,tkata}} \right) + A \right)} \quad (5-26)$$

$$A = \frac{V_{f,tkata}}{V_{r,tkata} K_{eq,tkata}} \cdot (K_{m,F6P,tkata} GAP + K_{m,GAP,tkata} F6P + GAP \cdot F6P)$$

Nukleotid-Synthese

$$v_{rppk} = \frac{V_{f,rppk} C5P^{n_{rppk,1}}}{K_{m,C5P,rppk}^{n_{rppk,1}} + C5P^{n_{rppk,1}}} \cdot \left(\frac{ADP}{K_{a,ADP,rppk}} \right)^{n_{rppk,2}} \quad (5-27)$$

DAHP-Synthase

$$v_{dahps} = \frac{V_{f,dahps} C5P \cdot PEP}{(K_{m,C5P,dahps} + C5P) (K_{m,PEP,dahps} + PEP)} \quad (5-28)$$

In Gleichung (5–10) wurde z.B. die bekannte Goldbeter-Kinetik (Goldbeter und Lefever 1972) für die von dem Enzym Phosphofructokinase katalysierte Phosphorylierung von F6P gewählt. Zusätzlich wurde die Kinetik multiplikativ mit Termen für das zweite Edukt ATP für zwei bekannte Aktivatoren F6P und ADP (Blangy et al. 1968; Torres und Babul 1991; Johnson und Reinhardt 1994) und für einen Inhibitor PEP (Deville-Bonne und Garel 1992; Johnson und Reinhart 1997) erweitert. Neben der Phosphofructokinase ist auch die Pyruvat-Kinase ein hochreguliertes Enzym in der Glykolyse. Die Reversibilität der Reaktion und die beiden benötigten Kofaktoren ADP bzw. ATP wurden über die Wahl einer reversiblen Ordered-Bi-Bi-Kinetik, s. Anhang D, als Grundform von Gleichung (5–20) berücksichtigt. Weiterhin wurden Terme für zwei bekannte Inhibitoren der Pyruvat-Kinase G6P und C5P stellvertretend für das eigentlich inhibierende Substrat Ri5P hinzugefügt (Malcovati und Valentini 1982). Im PPP ist hauptsächlich die 6PG-Dehydrogenase ein stark reguliertes Enzym, und wird vorwiegend durch glykolytische Intermediate beeinflusst (Silverberg und Dalziel 1973; Veronese et al. 1976; Orozco de Silva und Fraenkel 1979; Moritz et al. 2000), s. Gleichung (5–22) für eine entsprechende Wahl der Enzymkinetik.

Als Grundformen für die Kinetiken der meisten anderen Reaktionsschritte wurden einfache multiplikative Michaelis-Menten-Gleichungen, wie z.B. die Doppelsubstrat-Kinetik für die Pyruvat-Dehydrogenase, s. Gleichung (5–22), bzw. die zusammengezogene Trisubstrat-Kinetik für die Reaktion $v_{gap \rightarrow pep}$, s. Gleichung (5–18), oder die Hill-Gleichung für irreversible Einsubstrat-Kinetiken, s. Gleichung (5–16), eingesetzt. Da angenommen wurde, daß die Glucose-Aufnahme und die Umwandlung von G6P zu F6P unter den experimentellen Bedingungen schnell verläuft (Richter et al. 1975; Neijssel et al. 1977; Rohwer et al. 2000), wurden für diese Reaktionen sogar nur einfache Geschwindigkeitsgesetze n -ter Ordnung bzgl. der jeweiligen Reaktanden aufgestellt.

Die Parameter der kinetischen Gleichungen wurden vor der Optimierung auf Startwerte gesetzt, die unter *in-vitro* Bedingungen für den jeweiligen Parametertyp, d.h. K_m -, V_{max} -, K_i -Wert, größenordnungsmäßig gültig oder für die jeweilige Reaktion exakt bekannt sind. Nach der Optimierung resultierte eine von mehreren möglichen Wertekonstellationen für die Parameter, welche in Abschnitt 6.2.4 vorgestellt wird, s. Tab. 6–1 (S. 97).

Ergebnisse Einige Ergebnisse für die Simulation des in den Tabellen 5–3 und 5–4 vorgestellten Modells mit den optimalen Parameterwerten, s. Tab. 6–1, sind in den Abbildungen 5–6 und 5–7 zusammengefaßt.

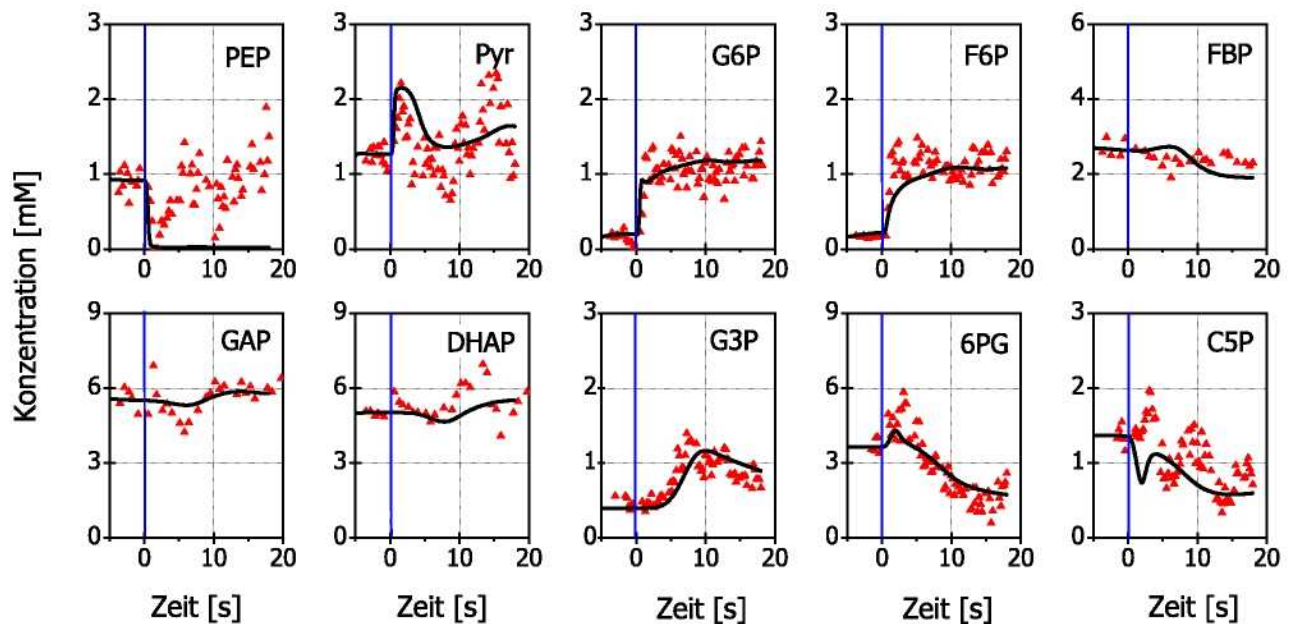


Abb. 5–6 Ergebnisse des Modells aus den Tabellen 5–3 und 5–4; simulierte Metabolittrajektorien

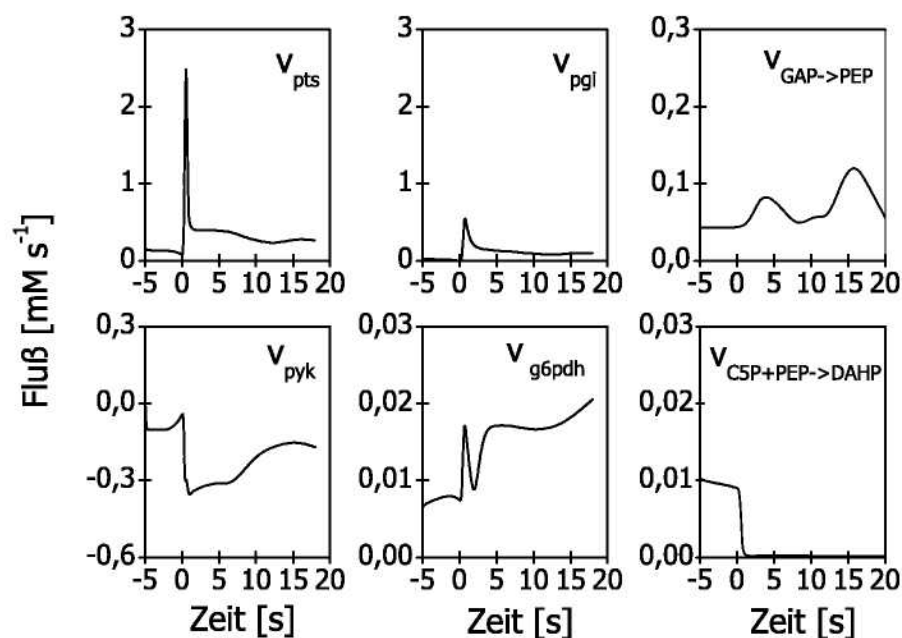


Abb. 5–7 Ergebnisse des Modells aus den Tabellen 5–3 und 5–4; simulierte Flüsse

Die Konzentrationswerte aller Metabolite können im Fließgleichgewichtszustand vor dem Substrat-Puls durch das Modell adäquat beschrieben werden, s. Abb. 5–6. Auch im Zeitraum nach dem Substrat-Puls zeigt sich für einen Großteil der Metabolite eine gute Übereinstimmung zwischen den simulierten und

den experimentellen Trajektorien. Insbesondere die Pyr-Anpassung gelang im Gegensatz zu (Chassagnole et al. 2002) recht gut. In Bezug auf die Intermediate PEP und C5P ist die Anpassungsgüte des Modells jedoch noch verbesserungswürdig. Denn, während der rasche Abfall der PEP-Konzentration infolge des Substrat-Pulses simulativ noch wiedergegeben werden kann, läßt sich langfristig kein Konzentrationsanstieg mehr feststellen, wie er experimentell beobachtet wurde. Selbst der negative Fluß über die Pyruvat-Kinase, d.h. eine überwiegende Rückreaktion von $\text{PEP} \rightarrow \text{Pyr}$, s. Abb. 5–7, und der nach dem Glucose-Puls ansteigende Fluß durch die Glykolyse reichen offenbar im Modell nicht aus, um die PEP-Konzentration wieder anzuheben. Evtl. tragen unter den vorliegenden experimentellen Bedingungen von Oxalacetat ausgehende anaplerotische Reaktionen, s. z.B. (Petersen 2000), in größerem Ausmaß zur Auffüllung der PEP-Konzentration bei.

Die simulierte Reaktion der C5P-Summenkonzentration auf den Substrat-Puls erfolgt zwar tendenziell richtig, da die Konzentration wie im Experiment abfällt, die Schwingungen werden jedoch nicht angemessen bzw. kurz nach dem Substrat-Puls sogar gegenteilig beschrieben. Als Ursache für die im Zeitraum von (0 - 5) s gegensätzlich verlaufende Schwingung konnte die abnehmende Konzentration des Kofaktors NADP der 6PG-Dehydrogenase, s. Anhang C.1, identifiziert werden, die zu dieser Zeit im Modell die Enzymaktivität limitiert. Daß die Dynamik der experimentell bestimmten C5P-Konzentration insgesamt weniger gut wiedergegeben werden konnte, wurde zunächst hauptsächlich auf die starken Vereinfachungen des Modells im Bereich dieses Stoffwechselabschnittes zurückgeführt. Dadurch können zeitweilige Konzentrationsverschiebungen zwischen den Metaboliten S7P, E4P und C5P bzw. regulative Beeinflussungen ihrer Reaktionen nicht berücksichtigt werden.

Das Flußprofil der Glucose-Aufnahmerate, s. Abb. 5–7 (links), stimmt in Gestalt und Größe fast exakt mit den Simulationsergebnissen von (Chassagnole et al. 2002; Sauter et al. 2002) für den gleichen *E.coli*-Stamm überein. Auch der Fluß in die Glykolyse über die Phosphoglucisomerase ist fast identisch mit Ergebnissen aus der Literatur (Chassagnole et al. 2002). Die Geschwindigkeit der von der G6P-Dehydrogenase katalysierten Eintrittsreaktion in den PPP ist jedoch um den Faktor 10 niedriger als in (Chassagnole et al. 2002) beschrieben. Dies liegt hauptsächlich darin begründet, daß das in Experiment WT1 beobachtete Abfallen der 6PG-Konzentration nur schwer unter Annahme eines hohen Flusses in den PPP simuliert werden kann, während in (Chassagnole et al. 2002) experimentell ein stetiger Anstieg der 6PG-Konzentration infolge des Pulses bestimmt wurde. Dieser läßt sich mit einem nach dem Puls nahezu unveränderten und bzgl. des Experimentes WT1 10-fach höheren Fluß in den PPP einfach erklären.

5.3.2 Zusammenfassung der Ergebnisse aller Modelle für das Experiment WT1

Insgesamt wurden 130 Modelle unter Postulierung von sechs verschiedenen Strukturen für das zugrundeliegende biochemische Reaktionsnetzwerk und von über 50 unterschiedlichen kinetischen Termen für die einzelnen Reaktionsgeschwindigkeiten erstellt, s. Anhang G. Von diesen 130 Modellen waren zwölf Modelle in der Lage, einen Großteil der experimentell bestimmten Metabolittrajektorien vor und nach dem Glucose-Puls adäquat wiederzugeben. Diese stimmen in der Zahl der insgesamt und der jeweils gut beschriebenen Metabolite mit relevanten veröffentlichten Modellen für Substrat-Puls-Experimente überein, s. Tab. 5–5. Lediglich die Zahl der DGL ist bei den aus der Literatur entnommenen Modellen höher. Das resultiert daraus, daß bei den letzteren auch DGL für experimentell nicht bestimmte Metabolite aufgestellt wurden, worauf bei den hier erstellten Modellen aus den oben genannten Gründen, s. S. 78, verzichtet wurde. Die beiden Modelle für den Organismus *S.cerevisiae* sind in Bezug auf die Zahl an DGL etwas komplexer, da die eukaryotische Hefezelle kompartmentiert ist, und daher cytosolische, mitochondriale und extrazelluläre Konzentrationen experimentell erfaßt und durch ein Modell wiedergegeben werden konnten. Dies entfällt für den Prokaryoten, *E.coli*, s. auch

den reduzierten Umfang des Modells aus (Chassagnole et al. 2002) im Vergleich mit (Visser et al. 2000; Theobald et al. 1997a).

Tab. 5–5 Charakteristika der Modelle zur Beschreibung des Experimentes WT1 im Vergleich mit Literaturwerten

Modell	DGL	Kinetiken	Para. ^a	Eff. ^b	Angepaßte intrazelluläre Metabolite ^c	
(Theobald et al. 1997a) ^d	21	21	84	10	G6P, F6P, FBP, GAP, PEP, Pyr , ATP, ADP , AMP, NADPH , Gluc _{ex} ^e , EtOH _{ex} , Acetat _{ex} , Gly _{ex}	(14)
(Visser et al. 2000) ^f	28	15	35	9	G6P, F6P , FBP, GAP, 3PG, PEP, Pyr Gluc _{ex} ^e , EtOH _{ex} , Acetat _{ex} , Gly _{ex}	(11)
(Chassagnole et al. 2002) ^g	18	30	117	13	G6P, F6P, FBP, G1P, GAP, PEP, Pyr , 6PG	(8)
Modell Nr. 1	10	17	101	8	G6P, F6P, FBP, GAP, DHAP, PEP, Pyr, 6PG , C5P	(9)
Modell Nr. 2	10	17	105	9	G6P, F6P, FBP , GAP, DHAP, PEP, Pyr, 6PG , C5P	(9)
Modell Nr. 3	10	17	102	7	G6P, F6P, FBP , GAP, DHAP, PEP, Pyr, 6PG, C5P	(9)
Modell Nr. 4	10	17	102	7	G6P, F6P, FBP , GAP, DHAP, PEP, Pyr, 6PG , C5P	(9)
Modell Nr. 5	10	17	102	7	G6P, F6P , FBP, GAP, DHAP, PEP, Pyr, 6PG, C5P	(9)
Modell Nr. 6	10	17	102	7	G6P, F6P, FBP , GAP, DHAP, PEP, Pyr, 6PG, C5P	(9)
Modell Nr. 7	10	15	91	12	G6P, F6P, FBP , GAP, DHAP, PEP , Pyr , 6PG, C5P	(9)
Modell Nr. 8	10	16	99	7	G6P, F6P, FBP , GAP, DHAP, PEP, Pyr, 6PG, C5P	(9)
Modell Nr. 9	9	16	92	7	G6P, F6P, FBP , GAP, DHAP, PEP, Pyr, 6PG, C5P	(9)
Modell Nr. 10	10	22	122	14	G6P, F6P, FBP, GAP, DHAP, G3P, PEP , Pyr, 6PG, C5P	(10)
Modell Nr. 11	10	23	131	15	G6P, F6P , FBP, GAP, DHAP, G3P, PEP , Pyr, 6PG, C5P	(10)
Modell Nr. 12	10	22	122	14	G6P, F6P, FBP , GAP, DHAP, G3P, PEP , Pyr, 6PG, C5P	(10)

^aPara: Zahl der Parameter.

^bEff.: Zahl der Effektoren.

^cFett-gedruckte Metabolite: geringere Übereinstimmung der simulierten Trajektorien mit den Meßdaten im Vergleich zu anderen simulierten Metaboliten desselben Modells. Kursiv gedruckte Metabolite: simulierte Trajektorien verlaufen relativ konstant, d.h. weisen kaum Schwingungen auf.

^dModell für *Saccharomyces cerevisiae* CBS 7336 (ATCC 32167).

^ec_{ex}: extrazelluläre Konzentration.

^fVereinfachtes Modell entwickelt aus dem in (Theobald et al. 1997a) dargestellten Modell.

^gModell für *Escherichia coli* K12.

Bei einem näheren Vergleich der im Rahmen dieser Arbeit erstellten zwölf Modelle, s. Anhang G.3, wird deutlich, daß diese auf nur zwei der insgesamt sechs postulierten Netzwerkstrukturen, d.h. den Strukturen III und V, s. Abb. G–1 und G–2 in Anhang G, basieren. Die anderen Strukturen erwiesen sich meist als zu stark vereinfacht, als daß auf ihnen beruhende Modelle eine größere Anzahl von experimentellen Trajektorien beschreiben könnten^{5,7}.

Im Gegensatz zu den Strukturen des Reaktionsnetzwerkes lassen sich keine gleichermaßen allgemein gültigen Aussagen für die Zuordnung der Kinetiken und Effektoren zu den Reaktionsschritten treffen. Denn die Freiheit für die Auswahl der Kinetiken ist enorm hoch. D.h. es ist nicht zwingend notwendig, für die einzelnen Reaktionsgeschwindigkeiten jeweils ganz bestimmte kinetische Terme mit zugehörigen, speziellen Effektoren auszuwählen, um eine hohe Wiedergabequalität der Meßdaten durch das Modell zu erreichen. Dies deutet sich bereits bei dem Vergleich der Zahlen der Effektoren, Kinetiken und Modellparameter an, die für die zwölf Modelle und für das ähnliche Literaturmodell, s. (Chassagnole et al. 2002), verwendet wurden^{5,8}.

^{5,7}Die Auswirkungen einer zu einfachen Struktur werden in Abschnitt 5.3.1 anhand der simulierten C5P-Konzentration diskutiert. Eine zu komplexe Struktur mit etlichen DGL und Modellparametern kann zu Schwierigkeiten bzgl. der Konvergenz des Optimierungsalgorithmus führen, s. Anhang G.

^{5,8}Gleiche Zahlen täuschen teilweise darüber hinweg, daß die Form der Kinetik oder die Art der Effektoren verschieden gewählt wurde, s. Anhang G.3 für eine detailliertere Auflistung der Modelle.

Auch die Modellaussagen, insbesondere bzgl. der simulierten Flußverteilung, unterscheiden sich. Damit ist gemeint, daß die Glucose-Aufnahmerate zwar noch von allen Modellen ähnlich prognostiziert wird, dann aber ausgehend vom Metaboliten G6P auf verschiedene Weise gesplittet wird. Dementsprechend lassen sich drei Klassen von Modellen unterscheiden:

1. die Modelle 1-7 (Struktur III), s. Abb. G-3 (S. 199):
es fließt eine größere Menge der aufgenommenen Glucose in den PPP als in die Glykolyse und aus dem PPP ausgehend von den C5-Körpern in die Nukleotid-Biosynthese ab.
2. die Modelle 8-9 (Struktur V), s. Abb. G-4 (S. 200):
der Fluß in den PPP ist auch hier höher als der in die Glykolyse; er wird aber durch einen hohen Abfluß in die AAA-Biosynthese kompensiert.
3. die Modelle 10-12 (Struktur V), s. Abb. G-5 (S. 201):
der Fluß in die Glykolyse ist höher als der Fluß in den PPP; ein hoher Abfluß aus dem PPP zweigt in diesen Modellen von den C5-Körpern in die Nukleotid-Biosynthese ab.

Bzgl. der simulierten Flußverteilung im PPP, die sich von der in (Vaseghi et al. 1999; Chassagnole et al. 2002) dargestellten Verteilung deutlich unterscheidet, sind die drei Klassen untereinander noch relativ ähnlich. Der ihnen allen gemeinsame hohe Abfluß aus dem PPP kann auf Schwierigkeiten bei der Anpassung der experimentellen 6PG- und C5P-Trajektorien zurückgeführt werden, die im Gegensatz zu den Experimenten in (Vaseghi et al. 1999; Chassagnole et al. 2002) vor dem Substrat-Puls vergleichsweise hohe Konzentrationswerte aufweisen und nach dem Puls stark abfallen. In Bezug auf die Aussagen der Klassen 2 und 3 für das Splitverhältnis des Flusses in die Glykolyse und den PPP ist jedoch in biologischer Hinsicht einer der deutlichsten Unterschiede bemerkbar, die überhaupt mit Modellen für diesen Abschnitt des Stoffwechsels erzielt werden können.

Damit lassen sich also offensichtlich insgesamt beliebig viele Modelle zur adäquaten Beschreibung eines Datensatzes konstruieren, deren Aussagen teilweise oder gar vollständig verschieden sein können. Weiterhin erweist sich bei der Erstellung solcher Modelle höchstens die zugrundeliegende Struktur des DGL-Systems als limitierend, während ansonsten eine nahezu freie Wahl der Kinetiken und Effektoren möglich ist, s. oben bzw. auch die Diskussion in (Theobald et al. 1997a; Chassagnole et al. 2002). Als wesentlicher Grund für die beiden Beobachtungen wird hauptsächlich die hohe Zahl an Parametern in den Modellen angesehen, die aus der Verwendung von parameterreichen *in-vitro* bestimmten Enzymkinetiken, anstelle z.B. der in (Visser et al. 2000) eingesetzten Linlog-Kinetiken, s. Tab. 5-5, resultiert. Dadurch sind die Modelle in der Form, wie sie bisher vorliegen, praktisch kaum mehr identifizierbar, was ihre Aussagekraft somit vermindert.

Aus diesem Grund war das Ziel weiterer theoretischer Untersuchungen, die hohe Parameterzahl auf die Kern-Parameter zu reduzieren, um klarere Aussagen über die Kinetiken bzw. Effektoren abzuleiten und die Modellfamilie über die Identifikation von fehlerhaften Modellstrukturen infolge der Parameterreduktion weiter einzugrenzen.

Zum Ausschluß bestimmter Modelle oder ganzer Modellklassen wären u.U. an dieser Stelle auch experimentelle Untersuchungen hilfreich gewesen. Ein weiteres Substrat-Puls-Experiment mit *E.coli* K12 konnte jedoch innerhalb des Zeitrahmens dieser Promotion nicht durchgeführt werden. Und es ist darüber hinaus fraglich, ob zur Eingrenzung der Modelle biochemisches Vorwissen, das meist aus *in-vitro*-Versuchen oder aus Experimenten unter Fließgleichgewichtsbedingungen entstammt, ohne weiteres auf die transienten Bedingungen der *in-vivo* Substrat-Puls-Experimente übertragen werden kann, vgl. z.B. den bei Glucose-limitierten *E.coli*-Zellen im Fließgleichgewichtszustand festgestellten hohen Fluß in den PPP (Edwards und Palsson 2000b).

6 Operationale Modellvalidierung

6.1 Einleitung

Um die Validität der Modellaussagen zu prüfen, werden in der Regel zusätzliche Datensätze benötigt, die unabhängig von den zur Modellerstellung verwendeten Daten sind. Solche Datensätze waren jedoch für die vorliegenden Modelle nicht erhältlich.

Infolgedessen wurde zunächst ein explorativer Ansatz zur operationalen Modellvalidierung gewählt, weil solche Ansätze nur das Modellverhalten untersuchen, ohne das Systemverhalten zum Vergleich heranzuziehen. Im Speziellen wurde die Sensitivitätsanalyse eingesetzt, s. Abschnitt 6.2, die auch insgesamt eines der am häufigsten verwendeten Verfahren darstellt (Page 1991; Saltelli et al. 2001). Denn bei ihr wird vor allem bestimmt, von welchen Parametern die Simulationsergebnisse am stärksten abhängen. Dadurch ist es möglich, wesentliche von evtl. zu vernachlässigenden Variablen zu trennen, sowie etwaige Strukturfehler im Modell aufzudecken. Dies war gerade das am Abschluß der Modellbildung formulierte Ziel für weitere Untersuchungen, d.h. die Parameterzahl der Modelle sollte reduziert und die Modellfamilie durch Falsifikation unzutreffender Modelle soweit wie möglich eingegrenzt werden, s. Abschnitt 5.3.2.

Darüber hinaus ist es auf der Basis der Sensitivitätsfunktionen möglich, die Fisher-Informationsmatrix zu bilden. Sie kann bei nichtlinearen Modellen u.a. als Grundlage für die Berechnung der Parameterschätzfehler genutzt werden, die wichtige Kenngrößen zur Bewertung eines Modells darstellen, s. Abschnitt 6.2.6.

Als letzte Option zur operationalen Modellvalidierung unter den gegebenen Bedingungen erschien sinnvoll, die Regressionsgüte der Modelle zwar noch in Bezug auf den gleichen Datensatz, aber mit anderen, anspruchsvolleren Kriterien als der Fehlerquadratsumme zu prüfen. Zu diesem Zweck wurde ein Verfahren getestet, das mit der sogenannten *Wild-Bootstrap*-Methode erzeugte Datensätze für einen statistischen Hypothesentest der Anpassungsgüte nutzt (Stute 1997; Stute et al. 1998), s. Abschnitt 6.3. Als Vorteile des Bootstrap-Verfahrens wurden vor allem angesehen, daß es im Gegensatz zu den meisten anderen Kriterien für die Beurteilung der Regressionsgüte die Meßunsicherheiten der Datenpunkte nicht benötigt^{6.1}, auch für Daten mit evtl. nicht normalverteilten Meßunsicherheiten verwendet werden kann und für die Reduktion einer Modellfamilie tauglich ist.

6.2 Sensitivitätsanalyse

6.2.1 Motivation

Die Auswirkungen, die die Änderung eines Modellparameters auf die Simulationsergebnisse hat, sind von Parameter zu Parameter verschieden. Beeinflussen schon kleine Änderungen des Parameters die simulierten Ausgabefunktionen entscheidend, so wird der Parameter als sensitiv bezeichnet und ist mit hoher Sicherheit auch für das reale System relevant, sofern das Modell Gültigkeit besitzt.

Auf der anderen Seite werden gerade Modelle mit einer hohen Anzahl von Parametern, auch insensitive Parameter beinhalten, deren Änderung das Modellergebnis relativ wenig beeinflusst, und die daher aus dem Modell eliminiert werden können (Turányi 1990; Smets et al. 2002). Die Parametersensitivitäten sind somit ein Maß dafür, wie stark das Modell auf die Änderungen des Parameters reagiert, d.h. ein Maß für die Wichtigkeit des Parameters.

^{6.1}Die Meßunsicherheiten für die Daten aus Experiment WT1 konnten z.B. aufgrund der fehlenden Dokumentation nicht bestimmt werden, s. Abschnitt 4.2.

Die Änderungen unterschiedlicher Modellparameter können u.U. die gleichen Auswirkungen auf die Ausgabefunktionen des Modells haben und sich somit gegenseitig kompensieren. In diesem Fall wird von korrelierenden Parametern gesprochen, und die Ähnlichkeit der Sensitivitäten stellt ein Maß für die Korrelation zwischen den Parametern dar.

6.2.2 Lokale Sensitivitätsfunktionen

Lokale Sensitivitätsfunktionen zeigen die Änderung der berechneten Lösungen eines Anfangswertproblems:

$$\frac{d\mathbf{c}(t, \mathbf{p})}{dt} = f(\mathbf{c}, \mathbf{p}), \quad \mathbf{c}(t_0) = \mathbf{c}_0, \quad (6-1)$$

für die Variation einzelner Modellparameter um gegebene Werte im Parameterraum. Die Auswirkung einer geringen Parameteränderung auf die Lösung kann in Form einer Taylor-Reihenentwicklung wiedergegeben werden (Saltelli et al. 2001):

$$\begin{aligned} c_j(t, \mathbf{p} + \Delta\mathbf{p}, \mathbf{c}_0) &= c_j(t, \mathbf{p}, \mathbf{c}_0) + \sum_{g=1}^q \frac{\partial c_j(t, \mathbf{p}, \mathbf{c}_0)}{\partial p_g} \Delta p_g \\ &+ \frac{1}{2} \sum_{g=1}^q \sum_{h=1}^q \frac{\partial^2 c_j(t, \mathbf{p}, \mathbf{c}_0)}{\partial p_g \partial p_h} \Delta p_g \Delta p_h + \dots, \end{aligned} \quad (6-2)$$

mit der Zeit t , dem m -dimensionalen Vektor der abhängigen Variablen $\mathbf{c}$, und dem q -dimensionalen Parametervektor $\mathbf{p}$. Die partiellen Ableitungen $\partial c_j(t, \mathbf{p}, \mathbf{c}_0)/\partial p_g$ heißen lokale Sensitivitäten erster bzw. $\partial^2 c_j(t, \mathbf{p}, \mathbf{c}_0)/\partial p_g \partial p_h$ lokale Sensitivitäten zweiter Ordnung, usw. Werden für eine lineare Annäherung die lokalen Sensitivitäten mit einer Ordnung größer als eins in Gleichung (6-2) vernachlässigt, ergibt sich durch Differentiation von Gleichung (6-1) nach dem Parametervektor $\mathbf{p}$ das folgende DGL-System für die Parametersensitivitäten:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{p}} \left[\frac{d\mathbf{c}(t, \mathbf{p}, \mathbf{c}_0)}{dt} \right] &= \mathbf{N} \cdot \frac{\partial \mathbf{v}(t, \mathbf{c}(t, \mathbf{p}, \mathbf{c}_0))}{\partial \mathbf{p}} \\ &= \mathbf{N} \cdot \left(\frac{\partial \mathbf{v}(\mathbf{p}, \mathbf{c}(t, \mathbf{p}, \mathbf{c}_0))}{\partial \mathbf{c}} \cdot \frac{\partial \mathbf{c}(t, \mathbf{p}, \mathbf{c}_0)}{\partial \mathbf{p}} + \frac{\partial \mathbf{v}(t, \mathbf{c}(t, \mathbf{p}, \mathbf{c}_0))}{\partial \mathbf{p}} \right) \end{aligned} \quad (6-3)$$

$$\frac{d}{dt} \left[\frac{\partial \mathbf{c}(t, \mathbf{p}, \mathbf{c}_0)}{\partial \mathbf{p}} \right] = \mathbf{N} \cdot (\mathbf{J}_c \cdot \mathbf{S} + \mathbf{J}_p) \quad (\text{Satz von Schwarz})$$

mit

$$\frac{\partial \mathbf{c}(t_0, \mathbf{p}, \mathbf{c}_0)}{\partial \mathbf{p}} = \mathbf{0}.$$

$\mathbf{J}_c$ wird als Jacobi-Matrix bzw. $\mathbf{J}_p$ gelegentlich als parametrische Jacobi-Matrix bezeichnet (Saltelli et al. 2001), und $\mathbf{S} = \left\{ \frac{\partial c_j(t, \mathbf{p}, \mathbf{c}_0)}{\partial p_g} \right\}$ stellt die Sensitivitätsmatrix dar. Das DGL-System für die Sensitivitäten bzgl. der Startkonzentrationen $\mathbf{c}_0$ läßt sich analog aufstellen, wobei $\partial \mathbf{c}(t_0, \mathbf{p}, \mathbf{c}_0)/\partial \mathbf{c}_0 = \mathbf{11}^T$ ist.

Direkte Methoden zur Berechnung der Sensitivitätsfunktionen basieren auf der numerischen Integration des DGL-Systems in Gleichung (6-3). Dazu müssen dem DGL-Löser zu jedem Zeitschritt der

Integration die aktuellen Werte der Jacobi-Matrizen $\mathbf{J}_c$ und $\mathbf{J}_p$ und somit auch die aktuellen Werte der abhängigen Variablen $\mathbf{c}(t, \mathbf{p}, \mathbf{c}_0)$ bekannt sein. Dies erfordert die vorherige (Atherton et al. 1975; Kramer et al. 1981) oder simultane Integration des Anfangswertproblems aus Gleichung (6–1), d.h. es ergibt sich ein neues Anfangswertproblem der Dimension: $m + m \cdot q$, mit m der Anzahl der abhängigen Variablen und q der Parameterzahl. Anfänglich wurden die Gleichungen (6–1) und (6–3) zwar simultan aber unabhängig voneinander gelöst, was sich als relativ rechenintensiv erweist (Dickinson und Gelinas 1976; Turányi 1990). Verbesserte Algorithmen, wie z.B. die DDM-Methode (decoupled direct method) (Dunker 1981; Dunker 1984), nutzen aus, daß Gleichung (6–1) und (6–3) dieselbe Jacobi-Matrix besitzen und numerische Integratoren für steife DGL somit die gleiche Schrittweite und Konsistenzordnung zur Lösung beider Gleichungen verwenden. Dadurch ist in jedem Zeitschritt nur eine einmalige Dekomposition der Jacobi-Matrix nötig, mit der zunächst Gleichung (6–1) und danach Gleichung (6–3) für alle Parameter gelöst wird.

Indirekte bzw. Finite-Differenzen-Methoden nähern die Lösung für die Elemente der Sensitivitätsmatrix z.B. durch die Verwendung von zentralen Differenzenquotienten an (Dunker 1981; Dunker 1984; Pandis und Seinfeld 1989):

$$\frac{\partial c_j(t, \mathbf{p}, \mathbf{c}_0)}{\partial p_g} \approx \frac{c_j(t, p_g + \Delta p_g, p_{h=1\dots q, h \neq g}, \mathbf{c}_0) - c_j(t, p_g - \Delta p_g, p_{h=1\dots q, h \neq g}, \mathbf{c}_0)}{2\Delta p_g}. \quad (6-4)$$

Bei einem q -dimensionalen Parametervektor sind also $2q$ Simulationen des Anfangswertproblems aus Gleichung (6–1) mit jeweils einem veränderten Parameterwert notwendig. Die Parameteränderung sollte dabei weder zu groß gewählt werden, um die Annahme der lokalen Linearität nicht zu verletzen, noch zu klein sein, um Rundungsfehler zu vermeiden (Rabitz et al. 1983). Weitere Methoden zur Bestimmung der Sensitivitätsfunktionen stellen z.B. die Green-Funktions-Methode (Hwang et al. 1978) oder die polynomische Approximationsmethode (Hwang 1983) dar.

Nach ihrer Berechnung können die lokalen Sensitivitätsfunktionen für eine Reihe von Zwecken eingesetzt werden:

- Modellreduktion: der Einsatz der lokalen Sensitivitätsfunktionen zur Modellreduktion wird in den Abschnitten 6.2.4 und 6.2.5 diskutiert;
- Abschätzung der Parameterschätzfehler: die Schätzfehler der Parameter werden beispielhaft in Abschnitt 6.2.6 berechnet;
- Parameteranpassung: a) die bevorzugte Änderung der jeweils hoch-sensitiven Parameter kann die Konvergenzgeschwindigkeit eines Optimierungsalgorithmus erhöhen, s. z.B. die in Anhang F.2 beschriebenen Gradienten-Verfahren; b) bei einem schlecht konditionierten Anpassungsproblem – d.h. die Daten liefern nicht genügend Information, um alle Parameter zu schätzen – ist es möglich, über eine Eigenvektor-Eigenwert-Analyse der gewichteten Kosensitivitätsmatrix schwer bestimm-bare sowie voneinander abhängige Parameter zu identifizieren (Turányi 1990), s. Abschnitt 6.2.5; sie sollten dann bei der Optimierung auf ihre nominalen Werte festgesetzt werden, soweit diese bekannt sind;
- Versuchsplanung: auf der Basis von lokalen Sensitivitätsfunktionen, wie z.B. der u.a. aus ihnen zusammengesetzten Fisher-Informationsmatrix (Ljung und Glad 1994a), können verschiedene Kriterien zur Vorhersage optimaler Versuchsbedingungen abgeleitet werden, bei denen z.B. die Modellparameter besser identifizierbar sind, s. z.B. (Munack und Posten 1989; Baltes et al. 1994; Versyck et al. 1997);

- Stabilitätsanalyse: die Stabilität dynamischer Systeme z.B. hinsichtlich der Störung von Zustandsvariablen des Modells zum Zeitpunkt t_1 wird über eine Eigenwertanalyse der lokalen Sensitivitätsmatrix bzgl. der Anfangswerte: $\mathbf{K}(t, t_1) = \{\partial c_j(t, \mathbf{p}, \mathbf{c}_0)/\partial c_{j,0}(t_1)\}$, $\mathbf{K}(t_1, t_1) = \mathbf{11}^T$, $t \geq t_1$, untersucht (Larter et al. 1984; Hedges Jr. und Rabitz 1985; Chassagnole et al. 2002);
- Metabolische Kontrollanalyse: die über die Integration des DGL-Systems aus Gleichung (6–3) bestimmten Parametersensitivitäten entsprechen gerade den Konzentrationskontrollkoeffizienten, s. Gleichung (2–8) auf S. 16. Auf ähnliche Weise lassen sich auch die Flußkontrollkoeffizienten berechnen (Mauch et al. 1997).
- Unsicherheitsanalyse: die Wahrscheinlichkeitsdichtefunktion oder die Varianz der Modellausgaben wird aus den Wahrscheinlichkeitsdichtefunktionen oder Varianzen der experimentellen Parameter ermittelt, die aus einer statistischen Analyse der Meßdaten bekannt sind; meist werden hierfür globale Sensitivitätsanalysen, wie z.B. Monte-Carlo- (Stolarski et al. 1978) oder FAST-Methoden (FAST: Fourier Amplitude Sensitivity Test) (Falls et al. 1979), eingesetzt, aber eine erste Abschätzung ist auch mit lokalen Sensitivitätsfunktionen durchführbar (Atherton et al. 1975);

6.2.3 Implementierung der Sensitivitätsberechnung in MMT

```

0  function FUN_(Increment,ObereGrenze,AnfangsWertVektor,ParameterVektor,DGL-System)
1    Initialisierung
2    Prozent←Increment;
3    for alle Parameter (k)
4      StartParaVektor(k)←ParameterVektor(k);
5    Beginn Sensitivitätsberechnung
6    while (Prozent less than or equal ObereGrenze) do
7      for alle Parameter(k)
8        for alle Parameter (l)
9          ParameterVektor(l)←StartParaVektor(l);
10     ParameterVektor(k)←StartparaVektor(k)+prozent·StartParaVektor(k);
11     StartVektor←KOMBINIERE(AnfangsWertVektor,ParameterVektor);
12     for alle Simulationszeitpunkte (j)
13       KonzOben(j, i)←LSODE(DGLSystem, StartVektor, Optionen);
14     ...ab Zeile 8 analog wiederholen für die Erniedrigung des Parameters...
15     for alle Spezies (i)
16       for alle Zeitpunkte (j);
17       Sensivitaet(i,j)←(KonzOben(i,j)-KonzUnten(i,j))/(2·prozent·StartParaVektor(k));
18       AUSGABE(Prozent,Sensivitaet(i,j));
19     Prozent←Prozent+Increment;
20   end;
21 end;

```

Abb. 6–1 Implementation der Sensitivitätsberechnung für variierende Änderungen der Parameterwerte in MMT1; ohne die Zeilen 6, 19, 20 ergibt sich die ebenfalls implementierte Sensitivitätsberechnung für eine feste prozentuale Änderung aller Parameter; ObereGrenze und Increment sind die aus der MMT1-Oberfläche entnommenen Übergabeparameter

Die Berechnung der lokalen Sensitivitätsfunktionen wurde erstmalig in MMT implementiert. Dazu ist die den DGL-Löser, LSODE, aufrufende MMT-Routine „fun_c“ so abgewandelt worden, daß die Parametersensitivitäten über die indirekte Methode, s. Gleichung (6–4), ermittelt und in eine Datei ausgegeben werden können. Die Sensitivitätsberechnung läßt sich dabei mit einer fest vorgegebenen prozentualen Änderung aller Parameterwerte durchführen oder mit Parameteränderungen, die aus den in Abschnitt 6.2.2 erwähnten Gründen innerhalb vorher festgelegter Bereiche variieren, s. Abb. 6–1. Durch

die Anbindung der Sensitivitätsroutine an die MMT-Oberfläche ist es möglich, dort die entsprechenden Parameter für die beiden Verfahren einzustellen.

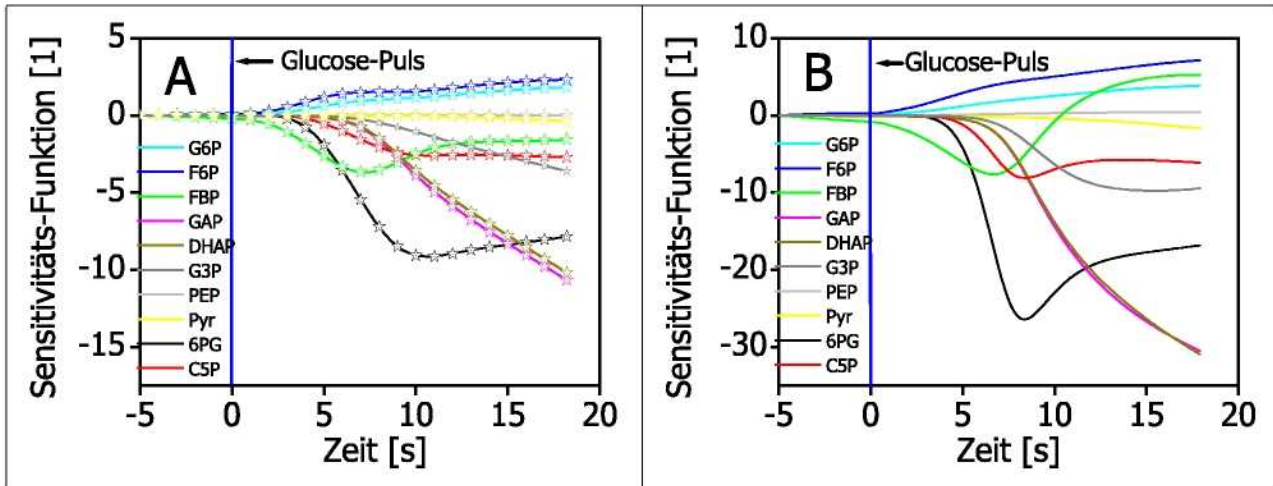


Abb. 6–2 Sensitivitätsfunktionen bzgl. des Parameters $K_{i,PEP,pfk}$ aus Modell 10 für Experiment WT1; A: kontinuierliche Linien: über die Finite-Differenzen-Methode, Sterne: über die direkte Methode berechnete Funktionen vor der Modellreduktion, B: über die Finite-Differenzen-Methode berechnete Funktionen nach der Modellreduktion, s. Abschnitt 6.2.4

Die Sensitivitätsfunktionen konnten ferner unter Verwendung einer in MMT2 im Nachhinein implementierten, direkten Methode (Haunschild und Wiechert 2003) verifiziert werden. Dabei stellte sich heraus, daß bei einer für die indirekte Methode optimierten relativen Änderung der Parameterwerte von $\Delta p_g/p_g = 10^{-5}$ die Ergebnisse beider Verfahren um höchstens 1% voneinander differieren, s. Abb. 6–2 (links). Daher wurde die Genauigkeit der über die Finite-Differenzen-Methode ermittelten Funktionen als ausreichend angesehen, um die Modellreduktion auf dieser Basis durchzuführen.

Zur automatischen graphischen Auswertung der von MMT ausgegebenen Datei mit den meist über 900 resultierenden Sensitivitätsfunktionen wurde eine C-Routine „sensplot.c“ implementiert, die das Programm XMGrace^{6.2} aufruft.

6.2.4 Eliminierung der insignifikanten Parameter

Da bei der Modellreduktion die unwichtigen Parameter über einen Vergleich der Parametersensitivitäten ermittelt werden, müssen letztere die gleiche physikalische Dimension besitzen oder dimensionslos sein. Aus diesem Grund wurden die Elemente der Sensitivitätsmatrix zunächst normiert (Dougherty et al. 1979):

$$\tilde{s}_{j,g}(t) = \frac{p_g}{c_j} \cdot \frac{\partial c_j(t, \mathbf{p}, \mathbf{c}_0)}{\partial p_g} = \frac{\partial \ln c_j(t, \mathbf{p}, \mathbf{c}_0)}{\partial \ln p_g}, \quad (6-5)$$

bevor sie weiteren Analysen unterzogen wurden. Weiterhin war eine Mittelung der normierten Sensitivitätsfunktionen über die Zeit und über die verschiedenen Zustandsvariablen notwendig, um den Gesamteinfluß eines Parameters auf den Modellausgang beurteilen zu können. Dazu wurden zwei verschiedene Methoden verwendet, die im Folgenden näher beschrieben werden, s. auch (Degenring et al. 2004a).

^{6.2}XMGrace 5.1.14, Grace Development Team, <http://plasma-gate.weizmann.ac.il/Grace/>

Methode 1 – Durchschnittliche Sensitivität Aus den lokalen Sensitivitätsfunktionen wurde für jeden Parameter ein Skalar berechnet, dessen Größe ein Schätzmaß für die Beeinflussung des Modellausgangs infolge der Eliminierung des korrespondierenden Parameters darstellt. In Analogie zu (Vajda et al. 1985) wurde zur Ermittlung des Skalars eine Objektivfunktion genutzt. Diese beschreibt die zeitlich gemittelte Änderung der Ausgabefunktionen eines Modells in Abhängigkeit von kleinen Parameteränderungen zur Anfangszeit der Simulation t_0 :

$$e(\Delta p_g) = \int_{t_0}^{t_n} \sum_{j=1}^m \left[\frac{c_j(t, p_{h=1\dots p, h \neq g}, p_g + \Delta p_g, \mathbf{c}_0) - c_j(t, \mathbf{p}, \mathbf{c}_0)}{c_j(t, \mathbf{p}, \mathbf{c}_0)} \right]^2 dt. \quad (6-6)$$

Die Sensitivität der Objektivfunktion bzgl. kleiner Parameteränderungen ergibt die durchschnittliche Sensitivität $os(p_g)$ und kann entweder direkt (Seigneur et al. 1982; Gautier et al. 1985) oder über die lokalen Sensitivitätskoeffizienten wie folgt berechnet werden:

$$os(p_g) = \frac{\partial e(\Delta p_g)}{\partial \ln p_g} = \sum_{i=1}^n \sum_{j=1}^m \left[\frac{\partial \ln c_{i,j}(t, \mathbf{p}, \mathbf{c}_0)}{\partial \ln p_g} \right]^2, \quad (6-7)$$

wobei hier das Zeitintegral in Gleichung (6-6) durch eine Summe über die Zeitpunkte $i = 1, \dots, n$ ersetzt wurde. $Os(p_g)$ ist eng verwandt mit der Fehlerquadratsumme, s. Gleichung (3-2), und wurde als Kriterium zur Sortierung der Parameter entsprechend ihrer geschätzten Reduktionsbedeutung gewählt.

Die Berechnung der durchschnittlichen Sensitivität und die Erstellung der Parameterrangliste wurde in die Routine „sensplot.c“ implementiert, s. Abschnitt 6.2.3.

Methode 2 – Hauptkomponentenanalyse der Kosensitivitätsmatrix In (Froemel 2003) wurde eine Methode zur Modellreduktion implementiert und angewendet, die eine Hauptkomponentenanalyse der normierten Kosensitivitätsmatrix vollzieht und die resultierenden Eigenvektoren auf drei unterschiedliche Arten auswertet, s. (Jolliffe 1972; Jolliffe 1972). Die Methode führt zwar eine etwas andere Mittelung der Sensitivitätsfunktionen über die Zeit und über die verschiedenen Spezies hinweg durch, sie schätzt aber ebenso wie die erste Methode eine Reihenfolge ab, in der die Parameter eliminiert werden sollten, s. (Froemel 2003).

Ergebnisse der Reduktion eines Stoffwechselmodells Entsprechend der nach Methode 1 erstellten Rangliste wurden beginnend bei dem Parameter mit der geringsten durchschnittlichen Sensitivität $os(p_g)$ alle Parameter der Reihe nach eliminiert, solange, bis sich deutliche Änderungen der Modellaussagen in Bezug auf die nach jedem Reduktionsschritt erneut simulierten Konzentrationen und Flüsse zeigten.

Dabei erwies es sich als effizient, die Parameter nicht, wie es mathematisch exakt wäre, gleich dem Neutralelement der jeweiligen Operation^{6.3} zu setzen, sondern sie entsprechend ihrer mechanistischen Bedeutung aus den Enzymkinetiken zu entfernen. Der Wert der maximalen Reaktionsgeschwindigkeit $V_{\max}$ ist z.B. immer multiplikativ mit dem restlichen Term verknüpft und meist die sensitivste Größe einer Enzymkinetik. Wird dieser Parameter als unwichtig für das Modell erkannt, weist dies darauf hin, daß die gesamte Kinetik eliminiert werden kann, d.h. $V_{\max}$ wird gleich Null und nicht gleich Eins gesetzt. Stellt als weiteres Beispiel die Konstante K_i eines multiplikativen Inhibitionsterms, s. z.B. Gleichung (5-4) (S. 81), eine insensitive Größe dar, so kann auch hier meist der ganze Inhibitionsterm aus dem Modell entfernt werden, was dann jedoch nicht gleichbedeutend mit der Aussage: $K_i = 0$, ist. Auf diesem Weg wurden pro Reduktionsschritt in der Regel nicht nur ein Parameter sondern mehrere

^{6.3}Das Neutralelement der Addition ist z.B. Null, das Neutralelement der Multiplikation bzw. der Division ist Eins, usw.

Parameter bzw. gesamte multiplikative Effektorterme oder Reaktionskinetiken aus dem Modell gelöscht. Die ursprünglich über die durchschnittliche Sensitivität abgeschätzte Reduktionsreihenfolge wurde infolge dieser Vorgehensweise nur geringfügig verändert, da Parameter, die bereits über die Elimination anderer Parameter wegfielen, in der Rangliste meist als nächst-insensitive Größen angezeigt wurden. Andererseits mußten jedoch einige Parameter mit einer relativ geringen durchschnittlichen Sensitivität in der Rangliste übersprungen werden, weil sie nicht aus dem Modell entfernt werden konnten. Dies ist bereits in (Vajda et al. 1985) beschrieben worden und läßt sich vermutlich darauf zurückführen, daß solche Parameter nur innerhalb kleiner Zeiträume oder für bestimmte Spezies signifikant sind und somit ihr Einfluß auf das Modell über die Bildung des Durchschnittswerts unterschätzt wurde.

Die Ergebnisse des gesamten Reduktionsprozesses werden im Folgenden anhand von Modell 10, s. Abschnitt 5.3.1, näher erläutert. Aus der Rangliste der Parameter geht hervor, daß nur acht der insgesamt 122 Parameter eine auffallend hohe durchschnittliche Sensitivität $os(p_g) \geq 99$ aufweisen. Sie lassen sich mit Ausnahme des Parameters $K_{m,6PG,pghl}$ der Glucose-Aufnahmereaktion bzw. glykolytischen Reaktionen zuordnen. Dabei hat die Inhibition der Phosphofructokinase durch PEP offenbar eine wichtige Bedeutung für das Modell, denn die betreffende Inhibitionskonstante besitzt die insgesamt größte durchschnittliche Sensitivität $os(K_{i,PEP,pfk}) = 214,6$ s. Abb. 6–2. Weitere 44 Parameter mit einer mittleren durchschnittlichen Sensitivität $10 \leq os(p_g) \leq 60$ aus vorwiegend glykolytischen Reaktionen erwiesen sich ebenfalls als signifikant. Für die verbleibenden 70 Parameter wurde eine geringe durchschnittliche Sensitivität $os(p_g) \leq 9$ berechnet, wobei sich 49 Parameter aus dem Modell entfernen ließen, ohne daß sich die Modellaussagen nennenswert veränderten, s. Abb. 6–3. Die übrigen 21 Parameter mußten in der Rangliste übersprungen werden, weil ihre Reduktion die simulierten Konzentrationen einiger Metabolite wenigstens innerhalb bestimmter Zeitabschnitte in einem höheren Ausmaß beeinflusste, s. die Diskussion im letzten Absatz.

Die eliminierten Parameter sind in Tab. 6–1 angegeben. Es zeigt sich, daß aufgrund der Streichung von Parametern in der Regel komplette Geschwindigkeitsterme, meistens für postulierte Abflüsse in die Biosynthese, wegfielen. Ansonsten ließen sich, wenn im originalen Modell bereits vereinfachte Enzymkinetiken verwendet wurden, nur noch multiplikative Effektorterme herauskürzen. Des weiteren konnten auch einzelne Parameter aus unverändert übernommenen *in-vitro*-Enzymkinetiken entfernt werden, wie z.B. bei den durch die Phosphofructokinase, die Pyruvat-Kinase oder die 6-Phosphogluconat-Dehydrogenase katalysierten Reaktionen. Daher wird es als vernünftig angesehen, künftig bereits bei der Modellerstellung für alle Reaktionsschritte Kinetiken mit möglichst wenigen Parametern einzusetzen.

Infolge der Parameterelimination wurde die Modellstruktur besonders im Stoffwechselabschnitt des PPP stark vereinfacht. Hier zeigte sich, daß sowohl die G6pdh-katalysierte als auch die kombinierte Transaldolase/Transketolase-Reaktion die Modellausgabe nur sehr gering beeinflussen. Denn, obwohl die beiden Reaktionen in dem vereinfachten Modell nicht mehr auftreten, sind die simulierten Konzentrationsverläufe fast identisch mit den Ergebnissen vor der Reduktion. Auch für die Flußverteilung ergeben sich keine wesentlichen Unterschiede, da die betreffenden Stoffflüsse bereits im Originalmodell relativ niedrig sind, s. Abb. 6–3. Damit ist es also möglich, die DGL der Metabolite des 6PG und C5P des PPP von dem übrigen DGL-System zu entkoppeln, so daß ein Teilmodell für die Glykolyse und ein weiteres zur Beschreibung des PPP resultiert, s. dazu das DGL-System des Ursprungsmodells in Tab. 5–3 (S. 79).

Insofern wurde durch die Sensitivitätsanalyse ein Strukturproblem des Modells 10 aufgedeckt, dessen Ursache vermutlich die hohe, experimentell bestimmte 6PG-Konzentration im Fließgleichgewichtszustand ist. Genährt durch den hohen Startwert kann die Fließgleichgewichtskonzentration von ca. 4 mM somit durch einen vernachlässigbar kleinen Fluß über die Glucose-6-Phosphat-Dehydrogenase

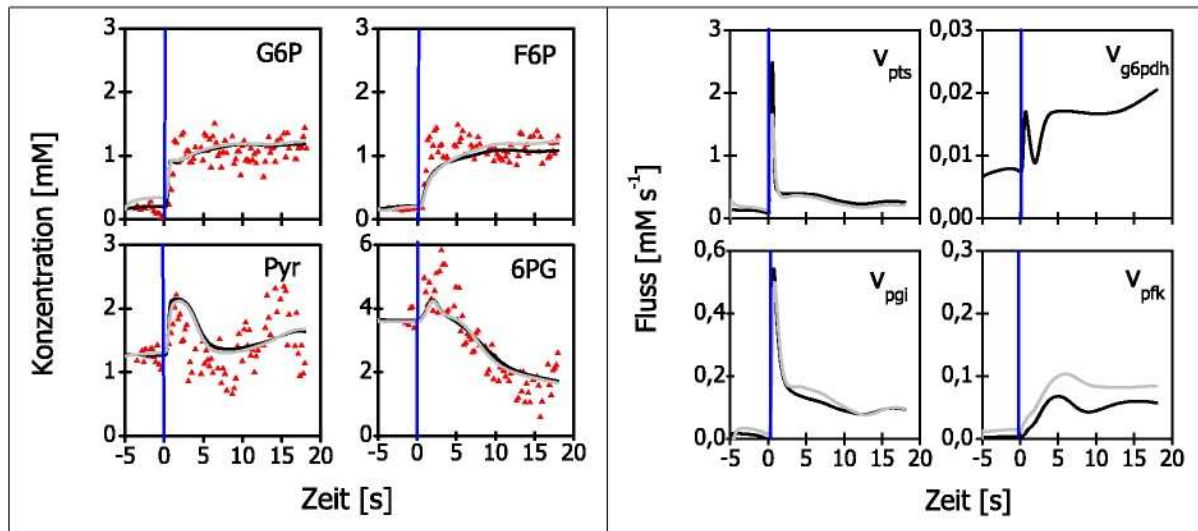


Abb. 6–3 Simulationsergebnisse von Modell 10; schwarze Trajektorien: vor der Parameterreduktion, graue Trajektorien: nach der Reduktion; links: Auswahl simulierter Konzentrationsverläufe, rechts: Auswahl simulierter Flüsse (die durch die G6pdh katalysierte Reaktion wurde infolge der Modellreduktion entfernt und ist daher nicht gezeigt)

aufrechterhalten werden. Nach dem Glucose-Puls wurde experimentell ein drastischer Abfall der 6PG-Konzentration bestimmt. Offenbar wirkt sich also in diesem Zeitraum der steile Anstieg der Konzentration des unmittelbaren Vorläuferstoffes G6P nicht auf den Metaboliten 6PG aus, so daß auch nach dem Puls ein nur geringer Fluß über die G6pdh simuliert werden konnte. Zur exakten Beschreibung des Abfalls der 6PG-Konzentration mußte zusätzlich sogar noch ein hoher Abfluß ausgehend von dem PPP-Intermediat angenommen werden, der eine essentielle Bedeutung innerhalb des Modells besitzt. Denn für den Parameter $K_{m,6PG,pghl}$ wurde der insgesamt viert-höchste Wert der durchschnittlichen Sensitivität bestimmt, und auch die anderen Parameter der Reaktionskinetik konnten nicht aus dem Modell eliminiert werden, s. Tab. 6–1. Die experimentell bestimmte Summe der Zucker-5-Phosphate weist die gleichen Charakteristika wie das 6GP-Profil auf, d.h. einen relativ hohen Konzentrationswert im Fließgleichgewichtszustand und einen Konzentrationsabfall nach dem Glucose-Puls. Die Übereinstimmung mit den anderen glykolytischen Intermediaten, F6P und GAP, ist dagegen weniger ausgeprägt. Daher kann im Modell die Dynamik der C5P-Konzentration allein durch den Fluß über die 6Pgdh-Reaktion beschrieben werden, ohne weitere Einflüsse aus der Glykolyse berücksichtigen zu müssen.

Die Beschreibung des PPP durch das Modell ist also letztendlich nicht adäquat. Das verbleibende Teilmodell für die Glykolyse ist jedoch realistisch und macht ähnliche Prognosen wie in (Chassagnole et al. 2002) beschrieben.

Tab. 6–1 Geschätzte Parameterwerte und korrespondierende Schätzfehler, s. Abschnitt 6.2.6, für Modell Nr. 10, s. Abschnitt 5.3.1, vor und nach der Modellreduktion^{6,4}; hochgestellte Ziffern in Spalte 4 bezeichnen Ergebnisse der Korrelationsanalyse, s. Abschnitt 6.2.5

Reaktions- kinetik	Parameter	Wert vor Reduktion	Wert nach Reduktion	Fehler	Dimension
v_{pts}	$k_{f,pts}$	4,358	4,4 ^{a1,40}	0,3	$\text{mM}^{-(n_{pts,1}+n_{pts,2}-1)}\text{s}^{-1}$
	$n_{pts,1}$	0,996	0,5 ^{a1,34}	0,1	1
	$n_{pts,2}$	1,053	0,6 ^{a1,34}	0,1	1
	$k_{r,pts}$	0,814	1,5 ^{b,34}	0,7	$\text{mM}^{-(n_{pts,3}+n_{pts,4}-1)}\text{s}^{-1}$
	$n_{pts,3}$	1,183	0,5 ^{a1,32}	0,3	1
	$n_{pts,4}$	0,141	–	–	1

Tab. 6–1 Geschätzte Parameterwerte und korrespondierende Schätzfehler, s. Abschnitt 6.2.6, für Modell Nr. 10, s. Abschnitt 5.3.1, vor und nach der Modellreduktion^{6,4}; hochgestellte Ziffern in Spalte 4 bezeichnen Ergebnisse der Korrelationsanalyse, s. Abschnitt 6.2.5

Reaktions- kinetik	Parameter	Wert vor Reduktion	Wert nach Reduktion	Fehler	Dimension
v_{pgi}	$k_{f,pgi}$	0,999	0,97 ^{c↑,44}	0,07	$\text{mM}^{-(n_{pgi,1}-1)}\text{s}^{-1}$
	$n_{pgi,1}$	0,952	1,3 ^{d↑,44}	0,2	1
	$k_{r,pgi}$	0,997	1,02 ^{c↓,44}	0,09	$\text{mM}^{-(n_{pgi,2}-1)}\text{s}^{-1}$
	$n_{pgi,2}$	0,996	1,0 ^{d↓,44}	0,2	1
v_{glp}	$V_{f,glp}$	1,976	1 ^{k↓,16}	$2 \cdot 10^3$	$\text{mM}^{(n_{glp,1}+1)}\text{s}^{-1}$
	$K_{m,G6P,glp}$	0,991	1 ^{k↑,7}	$5 \cdot 10^2$	mM
	$K_{m,ATP,glp}$	1,133	1 ^{k↑,16}	$6 \cdot 10^2$	mM
	$K_{i,G6P,glp}$	0,626	1 ^{k↑,7}	70	mM
	$n_{glp,1}$	0,769	1 ^{j↓,26}	70	1
	v_{pfk}	1,003	1 ^{o↑,0}	$3 \cdot 10^2$	1
v_{pfk}	D	0,999	1 ^{o↑,0}	$2 \cdot 10^2$	1
	e	0,999	1 ^{o↑,9}	10	1
	L	0,959	–	–	1
	c	0,998	–	–	1
	$K_{m,ATP,pfk}$	0,838	0,0 ^{o↓,17}	0,7	mM
	$n_{pfk,1}$	0,999	–	–	1
	$K_{i,ADP,pfk}$	0,999	–	–	mM
	$n_{pfk,2}$	0,981	–	–	1
	$K_{i,F6P,pfk}$	0,999	–	–	mM
	$n_{pfk,3}$	14,811	1 ^{o↑,2}	$2 \cdot 10^2$	1
	$K_{i,PEP,pfk}$	0,988	1 ^{o↓,61}	$3 \cdot 10^2$	mM
	v_{mur}	0,399	–	–	mM s^{-1}
	$K_{m,F6P,mur}$	1,786	–	–	mM
	$K_{m,ATP,mur}$	10,740	–	–	mM
v_{fbpase}	$V_{f,fbpase}$	0,124	–	–	mM s^{-1}
	$K_{m,fbpase}$	68,429	–	–	mM
	n_{fbpase}	1,960	–	–	1
v_{aldo}	$V_{max,aldo}$	0,897	1 ^{l↓,0}	$2 \cdot 10^3$	mM s^{-1}
	$n_{aldo,1}$	3,091	1 ^{l↑,17}	54	1
	$K_{m,F6P,aldo}$	4,991	1 ^{l↑,9}	$4 \cdot 10^2$	mM
	$K_{a,G3P,aldo}$	0,811	1 ^{l↑,0}	$1 \cdot 10^3$	mM
	$n_{aldo,2}$	2,671	3 ^{m,38}	4	1
$v_{fbp \rightarrow bm}$	$V_{f,fbp \rightarrow bm}$	0,044	–	–	mM s^{-1}
	$n_{fbp \rightarrow bm}$	5,040	–	–	1
	$K_{m,fbp \rightarrow bm}$	8,505	–	–	mM
v_{tim}	$V_{f,tim}$	1,557	1 ^{r↑,9}	$4 \cdot 10^2$	mM s^{-1}
	$K_{m,DHAP,tim}$	1,034	1 ^{r↑,9}	$5 \cdot 10^2$	mM
	$V_{r,tim}$	1,114	1 ^{r↓,16}	50	mM s^{-1}
	$K_{m,GAP,tim}$	0,834	1 ^{r↓,9}	$1 \cdot 10^3$	mM
v_{g3pdh}	$V_{f,g3pdh}$	1,559	1 ^{q↓,4}	$1 \cdot 10^3$	mM s^{-1}
	$n_{g3pdh,1}$	1,092	–	–	1
	$K_{m,DHAP,g3pdh}$	3,812	1 ^{q↑,9}	20	mM
	$K_{a,Cit,g3pdh}$	2,072	1 ^{q↑,5}	$6 \cdot 10^2$	mM
	$n_{g3pdh,2}$	5,692	6 ^{q↑,38}	1	1
v_{gly}	$V_{f,gly}$	2,947	1 ^{p↓,3}	$2 \cdot 10^3$	mM s^{-1}
	$n_{gly,1}$	1,592	2 ^{p↑,38}	2	1
	$K_{m,G3P,gly}$	1,662	1 ^{p↑,20}	44	mM
	$K_{a,F6P,gly}$	1,818	1 ^{p↑,3}	$1 \cdot 10^3$	mM
	$n_{gly,2}$	6,135	6 ^{p↑,38}	1	1
$v_{gap \rightarrow pep}$	$V_{f,gap \rightarrow pep}$	1,929	1 ^{e↑,19}	54	mM s^{-1}
	$K_{m,GAP,gap \rightarrow pep}$	3,672	1 ^{e↑,16}	$2 \cdot 10^2$	mM
	$K_{m,NAD,gap \rightarrow pep}$	4,290	1 ^{e↑,23}	17	mM
	$K_{m,ADP,gap \rightarrow pep}$	6,783	1 ^{e↓,23}	44	mM
v_{ser}	$V_{f,ser}$	0,285	–	–	mM s^{-1}
	n_{ser}	10,551	–	–	1
	$K_{m,GAP,ser}$	15,114	–	–	mM

Tab. 6–1 Geschätzte Parameterwerte und korrespondierende Schätzfehler, s. Abschnitt 6.2.6, für Modell Nr. 10, s. Abschnitt 5.3.1, vor und nach der Modellreduktion^{6,4}; hochgestellte Ziffern in Spalte 4 bezeichnen Ergebnisse der Korrelationsanalyse, s. Abschnitt 6.2.5

Reaktions- kinetik	Parameter	Wert vor Reduktion	Wert nach Reduktion	Fehler	Dimension
v_{pk}	$V_{f,pk}$	1,113	1 $j\downarrow,3$	$3 \cdot 10^3$	mM s^{-1}
	$K_{eq,pk}$	0,801	0,78 ^{s\downarrow,46}	0,05	1
	$K_{i,Pyr,pk}$	1,259	1 $s\uparrow,4$	$1 \cdot 10^3$	mM
	$K_{m,ADP,pk}$	1,010	1 $j\uparrow,6$	$6 \cdot 10^2$	mM
	$K_{i,PEP,pk}$	0,391	0 $s\uparrow,23$	8	mM
	$K_{m,PEP,pk}$	1,094	1 $j\uparrow,6$	$5 \cdot 10^2$	mM
	$V_{r,pk}$	2,444	1 $j\downarrow,3$	$1 \cdot 10^3$	mM s^{-1}
	$K_{m,ATP,pk}$	0,799	0 $j\uparrow,20$	4	mM
	$K_{m,Pyr,pk}$	0,107	–	–	mM
	$K_{i,ADP,pk}$	2,232	1 $j\downarrow,23$	71	mM
	$K_{i,G6P,pk}$	1,030	1 $j\downarrow,5$	$4 \cdot 10^2$	mM
	$n_{pk,1}$	2,074	2 $t\uparrow,31$	3	1
	$K_{i,C5P,pk}$	1,949	–	–	mM
	$n_{pk,2}$	0,089	–	–	1
$v_{pep \rightarrow bm}$	$V_{f,pep \rightarrow bm}$	1,321	–	–	mM s^{-1}
	$K_{i,PEP,pep \rightarrow bm}$	1,246	–	–	mM
	$K_{m,ATP,pep \rightarrow bm}$	41,119	–	–	mM
	$K_{m,PEP,pep \rightarrow bm}$	14,726	–	–	mM
v_{pdh}	$V_{f,pdh}$	0,996	1 $f\downarrow,23$	31	$\text{mM}^{n_{pdh}+1} \text{s}^{-1}$
	$K_{m,Pyr,pdh}$	0,996	1 $f\uparrow,23$	55	mM
	$K_{m,NAD,pdh}$	0,996	0,1 $f\uparrow,16$	0,4	mM
	$K_{i,AcCoA,pdh}$	0,998	1 $f\uparrow,31$	2	mM
	n_{pdh}	1,696	2 $f\uparrow,31$	2	1
v_{g6pdh}	$V_{f,g6pdh}$	8,384	–	–	mM s^{-1}
	$K_{m,NADP,g6pdh}$	3,533	–	–	mM
	$K_{i,G6P,g6pdh}$	0,727	–	–	mM
	$K_{eq,g6pdh}$	2,131	–	–	mM
	$K_{i,NADP,g6pdh}$	0,145	–	–	mM
	$K_{m,G6P,g6pdh}$	4,058	–	–	mM
v_{pgdh}	$V_{f,pgdh}$	0,283	1 $u\uparrow,3$	$5 \cdot 10^2$	mM s^{-1}
	$K_{eq,pgdh}$	1,712	1,81 $t,49$	0,02	mM^{-1}
	$K_{m,6PG,pgdh}$	26,696	–	–	mM
	$K_{m,NADP,pgdh}$	5,813	–	–	mM
	$V_{r,pgdh}$	1,130	1 $u\uparrow,3$	$1 \cdot 10^2$	mM s^{-1}
	$K_{m,C5P,pgdh}$	3,353	1 $u\downarrow,15$	$3 \cdot 10^2$	mM
	$K_{i,6PG,pgdh}$	1,388	1 $u\uparrow,15$	63	mM
	$K_{i,G6P,pgdh}$	11,543	9 $u\uparrow,3$	7	mM
	$n_{pgdh,1}$	3,268	2 $s\uparrow,22$	8	1
	$K_{i,F6P,pgdh}$	24,530	–	–	mM
	$n_{pgdh,2}$	54,790	–	–	mM
v_{pghl}	$V_{f,pghl}$	1,551	1 $g\downarrow,2$	$3 \cdot 10^2$	mM s^{-1}
	$n_{pghl,1}$	4,293	3,6 $g\downarrow,46$	0,3	1
	$K_{m,6PG,pghl}$	2,979	3,0 $g\uparrow,38$	0,8	mM
	$K_{a,F6P,pghl}$	1,135	1 $h,4$	45	mM
	$n_{pghl,2}$	7,404	12 $i,38$	1	1
v_{tkata}	$V_{f,tkata}$	1,134	–	–	mM s^{-1}
	$K_{eq,tkata}$	1,043	–	–	mM
	$K_{m,C5P,tkata}$	2,265	–	–	mM
	$K_{i,GAP,tkata}$	0,810	–	–	mM
	$V_{r,tkata}$	0,218	–	–	mM s^{-1}
	$K_{m,F6P,tkata}$	13,410	–	–	mM
v_{rppk}	$K_{m,GAP,tkata}$	0,868	–	–	mM
	$V_{f,rppk}$	0,308	–	–	mM s^{-1}
	$n_{rppk,1}$	1,477	–	–	1
	$K_{m,C5P}$	2,346	–	–	mM
	$K_{a,ADP}$	14,891	–	–	mM
	$n_{rppk,2}$	2,263	–	–	1

Tab. 6–1 Geschätzte Parameterwerte und korrespondierende Schätzfehler, s. Abschnitt 6.2.6, für Modell Nr. 10, s. Abschnitt 5.3.1, vor und nach der Modellreduktion^{6,4}; hochgestellte Ziffern in Spalte 4 bezeichnen Ergebnisse der Korrelationsanalyse, s. Abschnitt 6.2.5

Reaktions- kinetik	Parameter	Wert vor Reduktion	Wert nach Reduktion	Fehler	Dimension
v_{dahps}	$V_{f,dahps}$	0,400	1 $v_{1,9}$	10	mM s^{-1}
	$K_{m,C5P,dahps}$	5,591	1 $v_{1,9}$	32	mM
	$K_{m,PEP,dahps}$	6,238	1 $v_{1,9}$	$5 \cdot 10^2$	mM

Zusammenfassung der Ergebnisse für alle Modelle Die Ergebnisse für die Reduktion der anderen 11 in Tab. 5–5 dargestellten Modelle sind in Tab. 6–2 zusammengefaßt. Die Parameterzahl der Modelle konnte über die Methode der durchschnittlichen Sensitivität um ca. 55% reduziert werden, und ist damit in der Regel vergleichbar zu der Parameterzahl einer vereinfachten Variante des in (Theobald et al. 1997a) beschriebenen Stoffwechselmodells, s. (Visser et al. 2000) bzw. Tab. 5–5 (S. 85). Wie schon für das Modell 10 festgestellt wurde, führte die Streichung der Parameter bei allen Modellen hauptsächlich zu dem Verlust kompletter kinetischer Terme und zum Wegfall multiplikativer Effektortermine. Von ursprünglich unverändert eingesetzten *in-vitro* Enzymkinetiken blieben jeweils nur die $V_{\max}$ -, K_m -, K_i -Konstanten bzw. Parameter, die die allosterische Kooperativität der Enzyme ausdrücken, wie z.B. die Hill-Konstante n , erhalten. Parameter, die in Mehrsubstrat-Kinetiken den Einfluß von Mischtermen bzw. Kreuzprodukten beschreiben, wie z.B. die K_{iA} , K_{iB} , K_{iP} -Konstanten einer Bi-Bi-Kinetik, s. Anhang D, konnten dagegen meist eliminiert werden. Dies läßt sich evtl. damit begründen, daß im Gegensatz zu den *in-vitro* Versuchsbedingungen die Konzentration der Enzyme in der Zelle höher als die ihrer Substrate ist und somit die Bildung von ternären Komplexen, wie z.B. eines Enzym-Substrat₁-Substrat₂-Komplexes mit der Assoziationskonstante K_{iB} , seltener vorkommt. Insgesamt wird die obige Vermutung bestätigt, daß bei der Modellerstellung nicht die *in-vitro* bestimmten Enzymkinetiken, sondern vereinfachte Kinetiken zur Beschreibung der Umsetzungsgeschwindigkeiten eingesetzt werden sollten. Zu diesem Zweck werden vor allem die multiplikativen Kinetiken empfohlen, weil die in ihnen auftretenden $V_{\max}$ und K_m -Werte noch eine biochemische Bedeutung besitzen.

Bei einem Vergleich der beiden verwendeten Reduktionsmethoden fällt auf, daß mit dem ersten Verfahren, das die durchschnittliche Sensitivität der Parameter berücksichtigt, meist eine höhere Reduktionsrate erreicht wurde. Dies wird jedoch weniger auf das jeweils verwendete Kriterium als vielmehr auf die Vorgehensweise zur Reduktion zurückgeführt. Denn aus der über die Hauptkomponentenanalyse der Kosensitivitätsmatrix erstellten Parameterrangliste für Modell 11 geht hervor, daß die offensichtlich entfernbare G6pdh-katalysierte Reaktion erst angezeigt wird, nachdem andere, signifikante Parameter übersprungen wurden. Die Reduktion des Modells 11 wurde aber mit dem ersten Parameter, der nicht mehr aus dem Modell eliminiert werden konnte, abgebrochen, s. (Froemel 2003). Offensichtlich unterschätzt also auch das zweite Verfahren den Einfluß einiger Parameter, für die nur wenige Spezies innerhalb kleiner Zeiträume eine hohe Sensitivität aufweisen.

Im Anschluß an die Reduktion über die durchschnittliche Sensitivität der Parameter wurden die Modelle erneut an die Meßdaten angepaßt. Dabei zeigte sich, daß die resultierende Fehlerquadratsumme für die vereinfachten Modelle meist kleiner war als für die ursprünglichen Modelle. Dies wird in erster Linie einer verbesserten Ausgangssituation für den verwendeten Subplex-Algorithmus zugeschrieben, der als Gradienten-freies Optimierungsverfahren bei einer hohen Zahl von relativ insensitiven Modellparametern Konvergenzschwierigkeiten aufweist, s. auch die Diskussion auf S. 89.

^{6,4}Die Parameterwerte des ursprünglichen Modells können eigentlich nicht exakt bestimmt werden. Sie werden nur für den Vergleich mit den Werten des reduzierten Modells angegeben.

Tab. 6–2 Ergebnisse für die Parameterreduktion der Modelle aus Tab. 5–5; Schrägstriche auf der rechten Seite der Tabelle separieren die Ergebnisse der beiden Reduktionsmethoden: Durchschnittliche Sensitivität der Parameter / Hauptkomponentenanalyse der Kosensitivitätsmatrix bildet das Reduktionskriterium

Modell	DGL	vor der Reduktion				nach der Reduktion			
		Kin. ^a	Para. ^b	Eff. ^c	F [mM ²] ^d	Kin. ^a	Para. ^b	Eff. ^c	F [mM ²] ^d
Modell Nr. 1	10	17	101	8	0,194	11/ 15	44/ 86	4/ 6	0,169
Modell Nr. 2	10	17	105	9	0,165	7/ 12	34/ 59	3/ 5	0,120
Modell Nr. 3	10	17	102	7	0,132	6/ 12	26/ 56	2/ 4	0,115
Modell Nr. 4	10	17	102	7	0,142	6/ 13	26/ 67	2/ 3	0,114
Modell Nr. 5	10	17	102	7	0,130	6/ 11	26/ 49	2/ 3	0,109
Modell Nr. 6	10	17	102	7	0,105	7/ 12	34/ 56	3/ 3	0,105
Modell Nr. 7	10	15	91	12	0,075	14/ 14	79/ 77	11/ 9	0,085
Modell Nr. 8	10	16	99	7	0,187	7/ 12	29/ 68	2/ 5	0,175
Modell Nr. 9	9	16	92	7	0,115	12/ 10	64/ 57	5/ 1	0,115
Modell Nr. 10	10	22	122	14	0,110	14/ 14	73/ 79	9/10	0,150
Modell Nr. 11	10	23	131	15	0,109	14/ 20	64/101	9/10	0,086
Modell Nr. 12	10	22	122	14	0,151	12/ 18	73/104	9/13	0,155

^aKin.: Zahl der verwendeten kinetischen Terme.

^bPara.: Zahl der Parameter.

^cEff.: Zahl der Effektoren.

^dF: Fehlerquadratsumme für 2297 Datenpunkte, Gl.(3–2) wurde zusätzlich noch durch die Zahl der jeweils angepaßten Metabolite dividiert, in (Froemel 2003) ist *F* nicht dokumentiert und kann daher hier für die zweite Methode nicht angegeben werden.

Bei einer abschließenden Untersuchung der einzelnen Modellstrukturen wurde festgestellt, daß die Modelle 3-5 zu einer Grundstruktur vereinfacht werden konnten. Bei ihnen sowie bei den Modellen 1, 2, 6, 8, 9, 12 tritt darüber hinaus das gleiche Strukturproblem auf, d.h. sowohl die Aldolase- als auch die kombinierte Transaldolase / Transketolase-Reaktion konnte eliminiert werden. Dadurch wird die Glykolyse unterbrochen, und die Modelle sind somit biologisch bedeutungslos. Das Modell 11 besitzt hingegen ähnliche Strukturmängel, wie sie für das Modell 10 beschrieben worden sind. Denn die DGL für 6PG und C5P können auch hier vom übrigen DGL-System entkoppelt werden, während das verbleibende Teilmodell für die Glykolyse sinnvoll erscheint. Allein das Modell 7 behält nach der Reduktion eine insgesamt biologisch sinnvolle Modellstruktur bei. Bei ihm täuscht allerdings die geringe Fehlerquadratsumme, s. Tab. 6–2, darüber hinweg, daß nur vier Metabolite mit einer relativ hohen Konzentration, d.h. G6P, 6PG, DHAP und GAP, gut angepaßt werden konnten, während die sonstigen Intermediate eine nur sehr geringe Ähnlichkeit mit den Meßdaten aufwiesen.

Insgesamt konnten aufgrund der Sensitivitätsanalyse bei allen Modellen mit Ausnahme des Modells 7, das zur Wiedergabe der Meßdaten weniger geeignet ist, strukturelle Defizite im Bereich der glykolytischen Aldolase-Reaktion oder innerhalb des PPP festgestellt werden. Dies kann bedeuten, daß ein optimales Modell, das die Meßdaten hinreichend genau beschreibt und zugleich eine biologisch einwandfreie Modellstruktur besitzt, für diesen Datensatz noch nicht gefunden wurde. Angesichts der hohen Zahl an Modellen, die unter Voraussetzung der unterschiedlichsten Annahmen erstellt wurden, und in Anbetracht der Ergebnisse der Zeitreihenanalyse, s. Abschnitt 4.7.4 (S. 65), sowie des unphysiologisch hohen Wertes für die 6PG-Konzentration im Fließgleichgewichtszustand sollten allerdings auch die Meßdaten, insbesondere für die Metabolite 6PG und C5P aber auch für die glykolytischen Intermediate FBP, DHAP und GAP nochmals experimentell überprüft werden^{6,5}.

^{6,5}Die Schlußfolgerung für die glykolytischen Intermediate wurde in (Degenring et al. 2004a) noch nicht explizit gezogen, obwohl das Modellstrukturproblem schon bekannt war. Denn zu dieser Zeit stand die experimentelle Prüfung der Meßdaten für diese Metabolite noch aus.

In der Tat konnte bei einer im Anschluß an die Sensitivitätsanalyse durchgeführten Kontrolle der Meßdaten gezeigt werden^{6,6}, daß die ESI-LC-MS-Methode nach (Buchholz et al. 2001) zu hohe Konzentrationswerte für 6PG und C5P vortäuschte und die Konzentrationen dieser Metabolite eigentlich unterhalb der Nachweisgrenze lagen. Weiterhin wurde experimentell beobachtet^{6,7}, daß die den enzymatischen Assays zugesetzten Kofaktoren, d.h. vorwiegend die Stoffe NAD, NADH, NADP und NADPH, deren Umsatz zur Metabolitquantifizierung photometrisch verfolgt wird, sich bereits in größerem Umfang in der Probenlösung zersetzten, bevor die entsprechenden Enzyme den Assays zugegeben wurden^{6,7}. Somit liegt die Vermutung nahe, daß auch die Messungen der glykolytischen Intermediate aus Experiment WT1 von einem Meßfehler behaftet sind, der tendenziell zu hohe Meßwerte vorspiegelt, s. dazu (Kopperschläger und Augustin 1967; Kruger 1995).

Daher wurde im Folgenden darauf verzichtet, die Aussagen der bestehenden 12 Modelle für diesen Meßdatensatz, z.B. über die Berechnung von metabolischen Kontrollkoeffizienten, vertieft zu analysieren.

Wären nach Durchführung der in diesem Abschnitt vorgestellten Untersuchungen noch einwandfreie Modellstrukturen übrig geblieben, so hätten die Modelle mindestens noch auf das Vorhandensein von korrelierenden und damit nicht identifizierbaren Parametern geprüft und die Schätzfehler der Parameterwerte angegeben werden müssen, bevor ein solches Modell als adäquat zur Beschreibung der Stoffwechselvorgänge angesehen werden kann. Die beiden weiterhin zur operationalen Modellvalidierung notwendigen Arbeitsschritte werden im Folgenden aus Gründen der Transparenz beispielhaft anhand von Modell 10 gezeigt, da dieses Modell bereits detaillierter vorgestellt worden ist, s. Tab. 5–3, 5–4 und 6–1.

6.2.5 Zusammenfassen korrelierender Parameter

Einleitung Eine Hauptkomponentenanalyse der Kosensitivitätsmatrix wird in der Chemischen Kinetik eingesetzt (Vajda et al. 1985), um Reaktionsnetzwerke zu reduzieren, indem die wechselseitige Abhängigkeit zwischen den Modellparametern bestimmt wird. Voneinander abhängige, evtl. korrelierende Parameter können zu Gruppenparametern zusammengefaßt, bzw. Gruppen der Parameter von dem Modell ausgeschlossen werden.

Die Begründung, warum die Detektion voneinander abhängiger Parameter über die Eigenvektoranalyse möglich ist, läßt sich in der geometrischen Interpretation des folgenden, zur Fehlerquadratsumme analogen Gütefunktionalen finden:

$$Q(\hat{\mathbf{p}}) = \sum_{i=1}^n \sum_{j=1}^m [c_{ij}(\hat{\mathbf{p}} + \Delta\hat{\mathbf{p}}) - c_{ij}(\hat{\mathbf{p}})]^2, \quad (6-8)$$

Die nach dem zweiten Glied abgebrochene Taylor-Reihenentwicklung für Gleichung (6–8) im Punkt des geschätzten Parametervektors $\hat{\mathbf{p}}$ lautet:

$$Q(\hat{\mathbf{p}}) \approx \sum_{i=1}^n Q_i(\hat{\mathbf{p}}) + \mathbf{g}_i^T(\hat{\mathbf{p}})\Delta\hat{\mathbf{p}} + \frac{1}{2}(\Delta\hat{\mathbf{p}})^T \mathbf{H}_i(\hat{\mathbf{p}})\Delta\hat{\mathbf{p}}. \quad (6-9)$$

$\mathbf{g}_i(\hat{\mathbf{p}})$ ist der Gradientenvektor und $\mathbf{H}_i(\hat{\mathbf{p}})$ ist die Hesse-Matrix von $Q_i(\hat{\mathbf{p}})$ zum Zeitpunkt t_i an der Stelle $\hat{\mathbf{p}}$ mit $[\mathbf{g}_i(\hat{\mathbf{p}})]_g = \partial Q_i(\hat{\mathbf{p}})/\partial \hat{p}_g$ bzw. $[\mathbf{H}_i(\hat{\mathbf{p}})]_{gh} = \partial^2 Q_i(\hat{\mathbf{p}})/\partial \hat{p}_g \partial \hat{p}_h$. Da bei $\hat{\mathbf{p}}$ ein Minimum der

^{6,6}Mündliche Mitteilung, M. Oldiges, IBT2, Forschungszentrum Jülich, August 2002.

^{6,7}Mündliche Mitteilung, M. Oldiges, IBT2, Forschungszentrum Jülich, 11.4.2003.

Fehlerfläche vorliegt, gilt $Q_i(\hat{\mathbf{p}}) = 0, \mathbf{g}_i(\hat{\mathbf{p}}) = 0 \forall i$. Daraus folgt:

$$Q(\hat{\mathbf{p}}) \approx \frac{1}{2} \sum_{i=1}^n (\Delta \hat{\mathbf{p}})^T \mathbf{H}_i(\hat{\mathbf{p}}) \Delta \hat{\mathbf{p}}, \quad (6-10)$$

wobei die Hesse-Matrix durch den folgenden Ausdruck ersetzt werden kann:

$$\mathbf{H}_i(\hat{\mathbf{p}}) = \left[\frac{\partial^2 Q_i(\hat{\mathbf{p}})}{\partial \hat{p}_g \partial \hat{p}_h} \right]_{\substack{g=1, \dots, q \\ h=1, \dots, q}} = 2 \mathbf{S}_i(\hat{\mathbf{p}})^T \mathbf{S}_i(\hat{\mathbf{p}}) + \mathbf{Z}_i(\hat{\mathbf{p}}), \quad (6-11)$$

mit $[\mathbf{S}_i(\hat{\mathbf{p}})]_{jg} = \partial c_{ij}(\hat{\mathbf{p}}) / \partial p_g$. Der Term $\mathbf{Z}_i(\hat{\mathbf{p}})$ in Gleichung (6-11) kann vernachlässigt werden, da er zweite Ableitungen nach den Konzentrationen beinhaltet. Daraus ergibt sich eine Näherungsformel für Gleichung (6-8):

$$Q(\hat{\mathbf{p}}) \approx \bar{Q}(\hat{\mathbf{p}}) = (\delta \mathbf{p})^T \sum_{i=1}^n [\mathbf{S}_i(\hat{\mathbf{p}})^T \mathbf{S}_i(\hat{\mathbf{p}})] \delta \mathbf{p}, \quad (6-12)$$

für (infinitesimal) kleine Änderungen des Parametervektors $\delta \hat{\mathbf{p}}$. Gleichung (6-12) ist eine quadratische Funktion für die Veränderungen $\Delta \hat{p}_g, g = 1, \dots, q$, bzw. stellt ein Hyperparaboloid der Dimension $q+1$ dar. An dem festen Punkt ϵ definiert die Gleichung: $\bar{Q}(\hat{\mathbf{p}}) = \epsilon$, eine Höhenlinie der Fehlerfläche bzw. eine Hyperellipse der Dimension q im Parameterraum. Die Richtung und die Länge ihrer Hauptachsen kann über eine Hauptkomponentenanalyse der Kosensitivitätsmatrix bestimmt werden:

$$\sum_{i=1}^n \mathbf{S}_i(\hat{\mathbf{p}})^T \mathbf{S}_i(\hat{\mathbf{p}}) = \mathbf{\Gamma} \mathbf{\Lambda} \mathbf{\Gamma}^T. \quad (6-13)$$

$\mathbf{\Lambda}$ ist eine Diagonalmatrix, die die Eigenwerte von $\sum_{i=1}^n \mathbf{S}_i(\hat{\mathbf{p}})^T \mathbf{S}_i(\hat{\mathbf{p}})$ beinhaltet, $\mathbf{\Gamma}$ ist die zugehörige Matrix der normierten Eigenvektoren: $\gamma_g^T \gamma_g = \lambda_g$. Anschaulich bestimmen die Komponenten der Eigenvektoren die Richtung und die Eigenwerte die Länge der Hauptachsen der Hyperellipse. Werden die einzelnen Eigenvektoren, $\gamma_g^T, g = 1, \dots, q$, mit dem Parametervektor, $\hat{\mathbf{p}}$ multipliziert, so ergeben sich Linearkombinationen der ursprünglichen Parameter. Zu jeder der insgesamt q Linearkombinationen tragen jeweils nur die Parameter signifikant bei, deren korrespondierende Komponenten in dem betreffenden Eigenvektor einen betragsmäßig hohen Wert besitzen. Als signifikant wurden im Rahmen dieser Arbeit die Komponenten $[\tilde{\gamma}_h]_{o=1, \dots, \eta}$ des Eigenvektors, γ_h , angesehen, deren Beträge die untere Schranke:

$$u_g = \frac{a}{q} \sum_{g=1}^q \gamma_{g;h} \quad (6-14)$$

überschritten, wobei a auf 10 gesetzt wurde^{6.8}. Die simultane Änderung der $[\tilde{\gamma}_1]_{o=1, \dots, \eta}$ Komponenten des zu dem größten Eigenwert korrespondierenden Eigenvektor hat damit die insgesamt höchste Auswirkung auf die simulierten Konzentrationen, während die simultane Änderung der $[\tilde{\gamma}_q]_{o=1, \dots, \eta}$ Komponenten im Eigenvektor mit dem kleinsten Eigenwert die insgesamt geringste Änderung der Konzentrationen bewirkt. Da korrelierende Parameter die Ausgabefunktionen des Modells immer in ähnlicher Weise beeinflussen, bestehen die über die Hauptkomponentenanalyse gefundenen Linearkombinationen vorwiegend aus korrelierenden Parametern, s. auch (Vajda et al. 1985).

^{6.8} Willkürlicher festgelegter Wert, der sich im Rahmen dieser Arbeit bewährt hat.

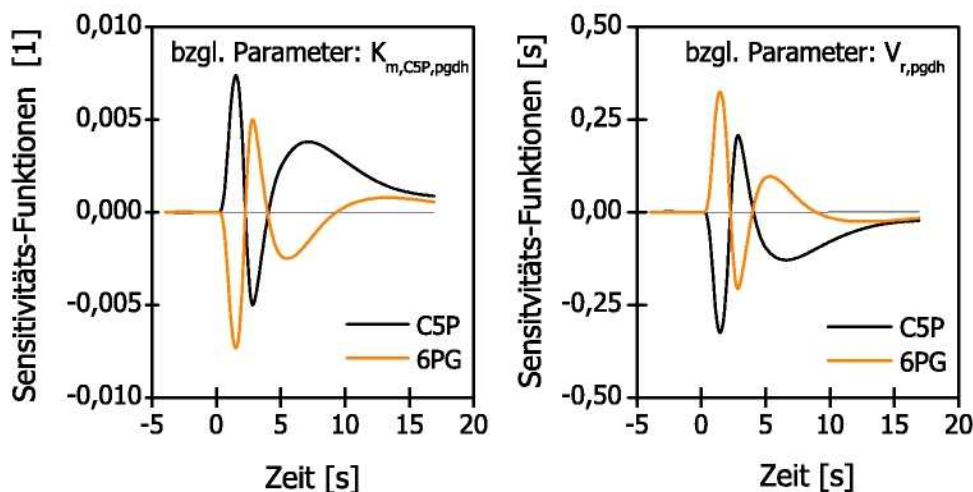


Abb. 6–4 Sensitivitätsfunktionen bzgl. zweier entgegengesetzt korrelierter Parameter aus Modell 10 (eine Vergrößerung des Parameters $K_{m,PEP,pgdh}$ bewirkt kurz nach dem Puls eine Erhöhung der C5P-Konzentration, während eine Vergrößerung von $V_{r,pgdh}$ eine Erniedrigung der C5P Konzentration nach dem Puls bewirkt); nur die Spezies 6PG und C5P reagieren sensitiv auf Änderungen der Parameter, die Parametersensitivitätsfunktionen der anderen Spezies besitzen über die gesamte Zeit hinweg einen Wert von ungefähr Null

Implementierung Zur Eigenvektoranalyse der Kosensitivitätsmatrix wurde die C-Routine „fim.c“ erstellt, die eine Datei mit den Sensitivitätsfunktionen im MMT2-Ausgabeformat einliest, nach Gleichung (6–5) normiert und die Funktion „nagf02“ der NAG C Bibliothek^{4,7} zur Berechnung der Eigenwerte bzw. Eigenvektoren aufruft. Ausgegeben werden die nach Größe sortierten Eigenwerte mit den korrespondierenden Eigenvektoren und eine $q \times q$ Matrix, in der die signifikanten Komponenten jedes Eigenvektors, s. Gleichung (6–14) mit einer Eins und die übrigen mit einer Null gekennzeichnet sind. Vor ihrem Einsatz wurde die Routine anhand des in (Vajda et al. 1985) dargestellten Beispielsystems der Wasserstoff-Brom-Reaktion verifiziert.

Ergebnisse In der vierten Spalte von Tab. 6–1 sind die Ergebnisse der Korrelationsanalyse für Modell 10 dargestellt. Die Zahlen 0-72 bezeichnen die Nummer des Eigenvektors, in dem die zu dem Parameter korrespondierende Komponente erstmalig die festgelegte untere Schranke aus Gleichung (6–14) überschreitet. Der Eigenvektor, der zu dem betragsmäßig größten Eigenwert gehört, besitzt dabei die Nummer 72. Zur Detektion der Korrelationen wurden beginnend mit dem zum kleinsten Eigenwert gehörigen Eigenvektor die Sensitivitätsfunktionen bzgl. solcher Parameter aufgetragen und auf Ähnlichkeit verglichen, welche den jeweils signifikanten Komponenten in den einzelnen Eigenvektoren entsprachen. Auf diese Weise ließen sich 22 Gruppen (Gruppen *a* bis *v*) untereinander korrelierter Parameter finden, wobei die korrelierenden Parameter in einer Gruppe die simulierten Trajektorien gleich- (↑) oder gegensinnig (↓) beeinflussen können. In Abb. 6–4 ist ein Beispiel von spiegelbildlich verlaufenden Sensitivitätsfunktionen bzgl. gegensinnig korrelierter Parameter gezeigt. Eine vollständige Korrelationsgruppe konnte meist über die Analyse von höchstens 10-15 aufeinanderfolgenden Eigenvektoren bestimmt werden. Ausnahmen ergaben sich nur, wenn die über die Zeit hinweg gemittelten, normierten Sensitivitätsfunktionen von Parametern aus einer Korrelationsgruppe sich deutlich in ihren Werten unterschieden.

Aus Tab. 6–1 geht weiterhin hervor, daß in der Regel fast alle Parameter einer Reaktionskinetik miteinander korrelieren und in Einzelfällen sogar die Parameter mehrerer Reaktionen voneinander abhängig sind, s. z.B. die Gruppen *j* und *s*. Damit sollte es möglich sein, das bereits durch die Elimination von

Parametern deutlich reduzierte Modell 10 über das Zusammenfassen von korrelierenden Parametern noch weiter zu vereinfachen. In ersten diesbezüglich durchgeführten Untersuchungen konnte das Modell 10 tatsächlich auf 32 Parameter reduziert werden, von denen nur noch maximal zwei einer der 22 gefundenen Korrelationsgruppen angehören. Für die Vereinfachungen wurden allerdings keine mathematischen Äquivalenzumformungen verwendet, sondern z.B. die Doppelsubstrat-Kinetik der DAHP-Synthase, s. Gleichung (5–4) (S. 82), auf die folgende Weise ersetzt:

$$v_{dahps} = \frac{V_{\max,dahps} \cdot \text{PEP} \cdot \text{C5P}}{(K_{m,\text{C5P},dahps} + \text{C5P}) \cdot (K_{m,\text{PEP},dahps} + \text{PEP})},$$

$$\Leftrightarrow v_{dahps} = \frac{V_{\max,dahps} \cdot \text{PEP} \cdot \text{C5P}}{K_{m,\text{PEP},dahps} K_{m,\text{C5P},dahps} + K_{m,\text{PEP},dahps} \cdot \text{C5P} + K_{m,\text{C5P},dahps} \cdot \text{PEP} + \text{C5P} \cdot \text{PEP}},$$

Annahmen: $K_{m,\text{PEP},dahps} K_{m,\text{C5P},dahps} \approx 0$; $\text{C5P} \cdot \text{PEP} \approx 0$; $K_{m,\text{PEP},dahps} \approx K_{m,\text{C5P},dahps}$;

$$\Rightarrow v_{dahps} \approx \frac{V'_{\max,dahps} \cdot \text{C5P} \cdot \text{PEP}}{\text{C5P} + \text{PEP}}, \quad (6-15)$$

und der optimale Wert für den Parameter $V'_{\max,dahps}$ über eine erneute Anpassung des Modells an die Meßdaten bestimmt. Da die Annahmen für die Vereinfachungen der Kinetiken nur in speziellen Fällen und auch nur näherungsweise gültig sind, sollen die detaillierten Reduktionsergebnisse hier nicht näher vorgestellt werden.

Als wesentliche generelle Resultate der Korrelationsanalyse für Modell 10 werden angesehen:

- daß über 50% Prozent der Parameter des bereits reduzierten Modells noch voneinander abhängig waren;
- daß das Zusammenfassen der Parameter nur über eine Neuformulierung des Modells gelang, wobei die grundsätzliche Form der Kinetik beibehalten werden mußte, s. z.B. die noch hyperbolische Funktion der vereinfachten Kinetik für die DAHP-Synthase in Gleichung (6–15)^{6,9};
- daß über die Neuformulierung des Modells nahezu alle korrelierenden Parameter eingespart werden konnten und somit das Modell nur noch 32 Parameter beinhaltet;

Damit scheint ein einzelner Datensatz aus einem Substrat-Puls-Experiment nicht auszureichen, um die Parameter von eingesetzten *in-vitro*-Stoffwechselkinetiken für die einzelnen Reaktionsschritte zu identifizieren, vgl. die hohe Parameterzahl von 122 im originalen Modell 10.

6.2.6 Parameterschätzfehler

Theoretische Grundlagen Die Schätzfehler der Modellparameter geben an, wie genau die Parameterwerte bei der Anpassung eines Modells an die Meßdaten bestimmt werden können. Sie sind wesentliche Kenngrößen für die Beurteilung eines Modells und sollten somit im Zuge der operationalen Modellvalidierung aus den lokalen Sensitivitätsfunktionen berechnet werden. Dazu wird von dem folgenden gebräuchlichen, aus der Likelihood-Funktion resultierenden Gütefunktional für die Bewertung des Gesamtfehlers zwischen den experimentellen Daten und den Ausgabefunktionen eines nichtlinearen Modells ausgegangen, s. (Ljung und Glad 1994a):

$$Q = \sum_{i=1}^n (\tilde{\mathbf{y}}_i - \mathbf{y}_i(t, \mathbf{y}_0, \hat{\mathbf{p}}))^T \mathbf{C}_i^{-1} (\tilde{\mathbf{y}}_i - \mathbf{y}_i(t, \mathbf{y}_0, \hat{\mathbf{p}})), \quad (6-16)$$

^{6,9}Bei Formulierung einer Kinetik erster Ordnung bzgl. beider Reaktanden wurde z.B. nur eine geringere Anpassungsgüte des Modells erreicht.

Gleichung (6–16) stellt die um die inverse Meßfehlerkovarianzmatrix:

$$\mathbf{C}_i = \delta_{jk} \text{var}(\tilde{\mathbf{y}}_i) = \begin{pmatrix} \text{var}(\tilde{y}_{1,i}) & \dots & 0 \\ 0 & \ddots & 0 \\ 0 & \dots & \text{var}(\tilde{y}_{m,i}) \end{pmatrix} \quad (6-17)$$

erweiterte Gleichung (6–8) dar. Denn die Vektoren $\mathbf{y}_i$ und $\tilde{\mathbf{y}}$ sind gleichbedeutend zu den hier vorliegenden Vektoren der intrazellulären Metabolitkonzentrationen. Damit kann das Gütefunktional in Analogie zu den Rechenoperationen (6–9) bis (6–12) näherungsweise durch den folgenden Ausdruck ersetzt werden:

$$Q \approx (\delta \mathbf{p})^T \underbrace{\sum_{i=1}^n [\mathbf{S}_i(t, \mathbf{c}_0, \hat{\mathbf{p}})^T \mathbf{C}^{-1} \mathbf{S}_i(t, \mathbf{c}_0, \hat{\mathbf{p}})]}_{\text{FIM}} \delta \mathbf{p} + n \cdot m. \quad (6-18)$$

Auch Gleichung (6–18) wird geometrisch als Hyperparaboloid der Dimension $q + 1$ interpretiert, wobei die Fehlerfläche diesmal aufgrund der Meßunsicherheiten ein von Null verschiedenes Minimum besitzt. Für $\delta \mathbf{p} > 0$ sind die Höhenlinien der Fehlerfläche Hyperellipsen, deren Koeffizientenmatrix die FIM ist. FIM bezeichnet die Fisher-Informationsmatrix des Schätzproblems (Ljung 1999). Sie ist die Inverse der SchätzfehlerKovarianzmatrix des besten erwartungstreuen Parameterschätzers, $\mathbf{C}_{\mathbf{p}}$, und beschreibt damit die bestmöglich erreichbare Parameterschätzgenauigkeit (Munack 1995). Im Fall des hier vorliegenden nichtlinearen Parameterschätzproblems liefert die Inverse der FIM die untere Schranke für die erreichbare Parameterschätzgenauigkeit:

$$\mathbf{C}_{\mathbf{p}} = \mathbb{E} \{ (\hat{\mathbf{p}} - \mathbf{p})(\hat{\mathbf{p}} - \mathbf{p})^T \} \geq \text{FIM}^{-1}. \quad (6-19)$$

Implementierung Da Gleichung (6–18) Gleichung (6–12) ähnlich ist, wurde die Berechnung der FIM ebenfalls in die Routine „fim.c“ implementiert. Dabei basiert die Schätzung der Meßfehlerkovarianzmatrix auf der Beziehung für die Meßunsicherheit aus Gleichung (6–30) mit $f_{\text{rel}} = u_g = 0,05 \text{ mol l}^{-1}$ für alle Metabolite und Zeitpunkte. Wird die Routine „fim.c“ dementsprechend zusätzlich mit der Angabe der Meßdatendatei aufgerufen, erfolgt die Ausgabe einer Datei „fim.ext“ mit der FIM, die in das Programm Matlab (Matlab 5.1.0.421, The Mathwork Inc., Copyright 1984-1997) geladen und invertiert werden kann.

Ergebnisse Die Quadratwurzeln der Hauptdiagonalelemente der Parameterschätzfehlerkovarianzmatrix wurden als Fehler der Parameterwerte in die fünfte Spalte von Tab. 6–1 eingetragen. Es zeigt sich, daß die Schätzfehler meist deutlich größer als die Werte der Parameter sind. Dies wird vorwiegend darauf zurückgeführt, daß die Parametersensitivitäten auch nach der Reduktion des Modells relativ kleine Werte besitzen, die meist im Zahlenbereich von: $-10^{-3} - +10^{-3}$, liegen. Auch die hier nicht aufgeführten Nebendiagonalelemente der ParameterschätzfehlerKovarianzmatrix sind teilweise groß. In ihnen spiegelt sich die noch vorhandene Korrelation zwischen einzelnen Modellparametern wieder.

Dies bestätigt nochmals die Aussagen des letzten Abschnitts, daß mit einem einzelnen Datensatz aus einem Substrat-Puls-Experiment höchstens ca. 30 Modellparameter identifiziert werden können.

6.2.7 Zusammenfassung

Lokale Sensitivitätsfunktionen können in vielerlei Hinsicht zur operationalen Modellvalidierung verwendet werden. Daher wurde die Berechnung der Funktionen im Rahmen dieser Arbeit in MMT implementiert, und die Parametersensitivitäten für die Modelle zur Beschreibung der Daten aus Experiment WT1, s. Abschnitt 5.3.2, bestimmt.

Verwendet wurden die Funktionen hier in erster Linie zur Elimination der insignifikanten Modellparameter auf der Basis eines Kriteriums, das auf den zeitlich und über die Spezies hinweg gemittelten Parametersensitivitäten beruht. Hierbei stellte sich heraus, daß insgesamt ca. 55% der Parameter aus den einzelnen Modellen eliminiert werden konnten. Im Anschluß an die Reduktion wiesen alle Modelle Strukturprobleme im Bereich des PPP bzw. der Aldolase-Reaktion auf, als deren Ursache experimentell fehlerhaft bestimmte Trajektorien der Metabolite 6PG, C5P, FBP, GAP und DHAP angesehen wurden. Dieser Umstand konnte über nachträglich durchgeführte experimentelle Untersuchungen bestätigt werden, s. (Oldiges 2004). Daher wurde auf eine vertiefte Analyse aller Modelle verzichtet und die Suche nach einem optimalen Modell zur Beschreibung der Daten nicht mehr weiter verfolgt.

Um die weiterhin notwendige Vorgehensweise zur operationalen Modellvalidierung wenigstens anzudeuten, wurde das bereits reduzierte Modell 10, s. Abschnitt 5.3.1, über die Hauptkomponentenanalyse der Kosensitivitätsmatrix auf voneinander abhängige Parameter untersucht. Dabei wurden 22 Gruppen untereinander korrelierender Parameter identifiziert, die jeweils fast alle Parameter einer Reaktionskinetik beinhalteten, d.h. maximal 3 Parameter des Modells waren pro Reaktionskinetik identifizierbar. Die Zusammenfassung der Modellparameter gelang nicht über mathematische Äquivalenzumformungen der Kinetiken. Es entstand durch die Vereinfachungen vielmehr ein neues Modell mit nur mehr 32 Parametern aus dem reduzierten Modell 10, das noch 73 Parameter beinhaltete.

Als Abschluß der Untersuchungen wurde die aus den Sensitivitätsfunktionen zusammengesetzte inverse Fisher-Informationsmatrix genutzt, um die Schätzfehler der Parameter des reduzierten Modells 10 zu bestimmen. Hierbei zeigte sich, daß selbst für die 73 Parameter des bereits über die Parametereliminierung reduzierten Modells 10 in der Regel noch große Schätzfehler resultierten. In Kombination mit der Korrelationsanalyse des Sensitivitätsfunktionen weisen diese Ergebnisse darauf hin, daß es auf der Basis eines Datensatzes aus einem Substrat-Puls-Experiment maximal gelingt, 30-35 Parameter zu identifizieren, s. auch die Diskussion in (Theobald et al. 1997a) bzw. (Visser et al. 2000).

Insofern sind *in-vitro* Kinetiken aufgrund ihrer hohen Parameterzahl für die Erstellung von Stoffwechselmodellen zur Beschreibung von Substrat-Puls-Experimenten wenig geeignet^{6,10}. Empfohlen wird daher die künftige Verwendung von Linlog-Kinetiken (Visser und Heijnen 2003) bzw. S-System-Ansätzen (Voit 1991) zur Beschreibung der Umsetzungsgeschwindigkeiten.

6.3 Bootstrap-Datensätze für den Test der Anpassungsgüte

6.3.1 Einleitung

Ziel der Untersuchungen in diesem Abschnitt ist es, verschiedene Kriterien dahingehend zu prüfen, ob sie als Maß für die Anpassungsgüte der vorliegenden DGL-Modelle ebenso gut oder besser geeignet sind als die bisher verwendete Fehlerquadratsumme.

Für Regressionsmodelle mit einer abhängigen Variablen $y = f(x_1, x_2, \dots, x_m; \mathbf{p})$ sind eine Reihe von verschiedenen Kriterien für die Beurteilung der Regressionsgüte entwickelt worden. Auch für solche

^{6,10}Es sei denn, es wurden fast alle Werte der Kinetiken über unabhängige Experimente unter *in-vivo*-Bedingungen bestimmt und können somit bei der Anpassung des Modells an einen bestimmten Meßdatensatz auf ihren Nominalwert festgesetzt werden.

Modelle stellt die Summe der quadrierten Differenzen zwischen den einzelnen Meßpunkten und den korrespondierenden berechneten Punkten ein sehr einfaches Kriterium dar.

Die Fehlerquadratsumme ist ein Spezialfall des anspruchsvolleren Kriteriums der Likelihood-Funktion bzw. der Log-Likelihood-Funktion (Fisher 1925; Norden 1972). Angewendet auf die Residuen, hängt diese Funktion zusätzlich von den Varianzen der Residuen ab, die evtl. an unterschiedlichen Stellen x_j verschiedene Werte besitzen, d.h. heteroskedastisch verteilt sind. Über die Berechnung dieses Kriterium kann damit in der Praxis berücksichtigt werden, daß Meßpunkte, die mit einer kleinen Meßunsicherheit erfaßt wurden, besser durch das Modell wiedergegeben werden sollten.

Auch die Zahl der Modellparameter ist eigentlich für die Bewertung der Anpassungsgüte relevant. Denn es ist einfacher, ein Modell mit relativ vielen Parametern an die Daten anzupassen, als die gleiche Wiedergabequalität mit einem Modell zu erreichen, das wenige Parameter beinhaltet, s. z.B. das Beispiel aus (Burnham und Anderson 1998, S. 22). Daher erweitert das Akaike-Information-Kriterium (AIK) den Wert der Log-Likelihood-Funktion um die Parameterzahl des Modells (Akaike 1973).

Neben der Berechnung von skalaren Zahlen zur Beurteilung der Regressionsgüte haben sich zu diesem Zweck auch Tests aus der statistischen Entscheidungstechnik bewährt, s. u.a. (Cox et al. 1988; Eubank und Spiegelman 1990; Su und Wei 1991; Buckley 1991; Eubank und Hart 1993; Härdle und Mammen 1993). Bei ihnen wird eine Nullhypothese: das Modell eignet sich zur Wiedergabe der Daten, gegen die Alternativhypothese geprüft. Die unterschiedlichen Kriterien zur Ablehnung des Modells basieren ebenfalls auf einem Maß für die Abweichung zwischen dem Modell und den evtl. auch geglätteten Meßdaten.

Der Vorteil des hier untersuchten Testverfahrens besteht darin, daß die Verteilungsfunktion der Residuen bzw. die evtl. heteroskedastische residuale Varianz nicht a-priori benötigt, sondern über die Erzeugung von künstlichen Bootstrap-Datensätzen abgeschätzt wird. Die Bestimmung der für die meisten anderen Kriterien, wie auch für die Likelihood-Funktion, benötigten Varianzen fällt nämlich im vorliegenden Fall schwer, da aus Zeitgründen die Substrat-Puls-Experimente nicht beliebig oft wiederholt werden können.

Die drei vorgestellten Kriterien wurden so erweitert, daß sie zur Beurteilung der Anpassungsgüte eines DGL-Modell mit mehreren abhängigen Variablen herangezogen werden können. Den Schwerpunkt der Untersuchungen bildete dabei das auf den Bootstrap-Datensätzen beruhende Testverfahren, weil es bisher noch nicht für biotechnologische Anwendungszwecke eingesetzt wurde.

6.3.2 Theoretische Grundlagen

Aus Gründen der Einfachheit werden die Grundlagen des Bootstrap-Näherungsverfahren zunächst für Regressionsmodelle mit jeweils einer abhängigen Variablen erklärt, für die das Verfahren ursprünglich hergeleitet wurde, s. (Stute 1997; Stute et al. 1998). Im Anschluß werden dann die vorgenommenen mathematischen Erweiterungen des Verfahrens dargestellt, die erforderlich waren, um die Methode auf die vorliegenden DGL-Modelle mit mehreren abhängigen Variablen anwenden zu können.

Eine Regressionsfunktion $m(\mathbf{x})$ drückt den Erwartungswert $\mathbb{E}$ einer beobachteten Größe $\tilde{y}$ in Abhängigkeit von dem Vektor der unabhängigen Variablen $\tilde{\mathbf{x}}$ aus:

$$m(\mathbf{x}) = \mathbb{E}[\tilde{y} \mid \tilde{\mathbf{x}} = \mathbf{x}], \quad \mathbf{x} \in \mathbb{R}^d. \quad (6-20)$$

Handelt es sich um ein parametrisches Regressionsmodell, so gehört m zu einer Familie von Funktionen $\mathcal{M} = \{m_{\mathbf{p}} : \mathbf{p} \in \mathcal{P}\}$, die durch den q -dimensionalen Vektor $\mathbf{p}$ parametrisiert werden. Dies ist

gleichbedeutend mit der Aussage:

$$m(\mathbf{x}) = m_{\mathbf{p}_0}(\mathbf{x}), \quad (6-21)$$

für einen unbekannten „wahren“ Parametervektor $\mathbf{p}_0$, der in der Praxis über die Anpassung des Modells an die Meßdaten auf $\hat{\mathbf{p}}$ geschätzt wird. Zur Prüfung der Regressionsgüte wird dann eine H_0 -Hypothese getestet: das Modell $m_{\hat{\mathbf{p}}}(\mathbf{x})$, dessen Parameter durch Anpassung an die Meßdaten bestimmt wurden, kann aufgrund seiner Anpassungsgüte mit einer gewissen Irrtumswahrscheinlichkeit α als gültig erachtet werden:

$$H_0 : m \in \mathcal{M} \quad \text{versus} \quad H_1 : m \notin \mathcal{M}, \quad (6-22)$$

Der hier angewendete Test für die H_0 -Hypothese basiert auf dem empirischen Prozeß^{6.11} der Residuen als Maß für die Ähnlichkeit zwischen dem Regressionsmodell und den Daten (Stute 1997):

$$\begin{aligned} R_n(\mathbf{x}) &= n^{1/2} \{I_n(\tilde{\mathbf{x}}) - \mathbb{E}[I_n(\tilde{\mathbf{x}}) \mid \tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n]\}, \\ &= n^{1/2} \sum_{i=1}^n 1_{\{\tilde{x}_i \leq \tilde{\mathbf{x}}\}} [\tilde{y}_i - m(\tilde{\mathbf{x}}_i)], \end{aligned} \quad (6-23)$$

wobei die Residuen durch $\epsilon_i = \tilde{y}_i - m(\tilde{\mathbf{x}}_i)$, $1 \leq i \leq n$, gegeben sind. Entsprechend dem Theorem 1.1 in (Stute 1997) geht gemäß dem Gesetz der großen Zahlen der empirische Prozeß der Residuen für eine hohe Zahl an Stichproben in einen Grenzprozeß $R_\infty(\mathbf{x})$ mit einer $N(0, \sigma^2(\mathbf{x}))$ -Normalverteilung über. Das liefert das Kriterium für den Test der H_0 -Hypothese. Dazu wird $R_n^1(\mathbf{x})$ für das zu testende Modell ermittelt, d.h. m in Gleichung (6-23) durch $m_{\hat{\mathbf{p}}}$ ersetzt^{6.12}, und die H_0 -Hypothese verworfen, wenn eine aus $R_n^1(\mathbf{x})$ berechnete Prüfgröße einen korrespondierenden kritischen Wert des Grenzprozesses überschreitet.

Für die Durchführung des Tests wird damit zunächst eine Näherung der $N(0, \sigma^2(\mathbf{x}))$ -Verteilungsfunktion von R_n^1 benötigt. Weiterhin muß eine geeignete Prüfgröße und ein geeigneter kritischer Wert definiert werden.

Die $N(0, \sigma^2(\mathbf{x}))$ -Verteilungsfunktion wird durch die Verteilung einer großen Zahl von $R_n^*(\mathbf{x})$ -Funktionen angenähert, die aus geeigneten künstlichen Datensätzen $(\mathbf{x}_i^*, y_i^*)$ berechnet werden. Es ist beweisbar (Stute et al. 1998), daß eine konsistente Näherung für die Verteilung der $R_n(\mathbf{x})$ -Funktion mit Daten erreicht wird, die über die sogenannte „Wild-Bootstrap“-Methode generiert werden (Efron 1979; Wu 1986; Liu 1988; Mammen 1992):

$$\begin{aligned} \tilde{\mathbf{x}}_i^* &= \tilde{\mathbf{x}}_i & \epsilon_i^* &= \bar{\epsilon}_i^* V_i^* \\ \tilde{y}_i^* &= m_{\hat{\mathbf{p}}}(\tilde{\mathbf{x}}_i) + \epsilon_i^* & \bar{\epsilon}_i^* &= \tilde{y}_i - m_{\hat{\mathbf{p}}}(\tilde{\mathbf{x}}_i) \\ & & V_i^* &\sim N(0, 1). \end{aligned} \quad (6-24)$$

^{6.11}Die Abbildung $\hat{F}_n: \mathbb{R} \times \Omega \rightarrow [0, 1]$ mit $\hat{F}_n(x, \omega) = (\#\{i : 1 \leq i \leq n, x_i(\omega) \leq x\})/n$ heißt empirische Verteilungsfunktion der Zufallsstichprobe $(x_1, \dots, x_n)$ und gibt die relative Häufigkeit der Stichprobenwerte wieder, die den Schwellenwert x nicht überschreiten. Diese Abbildung kann als Familie $\{\hat{F}_n(x), x \in \mathbb{R}\}$ von Zufallsvariablen $\hat{F}_n(x) : \Omega \rightarrow [0, 1]$ aufgefaßt werden, welche als empirischer Prozeß bezeichnet wird.

^{6.12}Theorem 1.2 in (Stute 1997) bestätigt, daß auch die $R_n^1(\mathbf{x})$ -Funktion für das an die Daten angepaßte Modell $m_{\hat{\mathbf{p}}}$ für eine große Zahl an Stichproben sich einem Grenzprozeß mit Normalverteilung annähert. Der obere Index 1 steht für den residualen empirischen Prozeß, der für das angepaßte Modell und den originalen Datensatz berechnet wurde.

Als Prüfgröße eignet sich z.B. die Kolmogorov-Smirnov (K.S.)-Teststatistik (Kolmogorov 1941; Smirnov 1948):

$$D_n^1 = \sup_{\mathbf{x}} |R_n^1(\tilde{\mathbf{x}}_i)|, \quad (6-25)$$

oder die Cramér-von Mises (CvM)-Teststatistik, s. (Stute 1997):

$$[W_n^1]^2 = \int_{\mathbb{R}^d} [R_n^1(\mathbf{x})]^2 F_{1n}(d\mathbf{x}) \quad \text{bzw. diskret:} \quad [W_n^1]^2 = n^{-1} \sum_{i=1}^n [R_n^1(\tilde{\mathbf{x}}_i)]^2. \quad (6-26)$$

Die kritischen Werte $c_{1-\alpha}$ werden entsprechend als die $b \cdot (1 - \alpha)$ ten Anordnungswerte der K.S. bzw. CvM-Teststatistiken definiert, die aus den R_n^* -Funktionen von b Bootstrap-Datensätzen berechnet wurden.

Überschreiten die aus den R_n^1 -Funktionen ermittelten Teststatistiken die kritischen Werte, so muß das Modell mit einer vor dem Testbeginn festgelegten (Irrtums)-Wahrscheinlichkeit α abgelehnt werden.

Um dieses Hypothesentestverfahren von Regressionsmodellen auf die hier vorliegenden Lösungen der gewöhnlichen DGL-Systeme erster Ordnung zu übertragen, wurden zwei Änderungen vorgenommen:

- der für den multivariaten Fall definierte Vektor der Regressoren $\tilde{\mathbf{x}}_i$, $1 \leq i \leq n$ wurde durch die unabhängige Variable, die Zeit, ersetzt, da die abhängigen Variablen, die Metabolitkonzentrationen, zu $1 \dots i \dots n$ voneinander unabhängigen Zeitpunkten t_i beobachtet wurden.
- um das Verfahren auf Mehrgleichungssysteme anwenden zu können, wurde unter der Annahme, daß die Residuen verschiedener abhängiger Variablen nicht korreliert sind, für jede abhängige Variable $1 \leq j \leq m$ der empirische Prozeß der Residuen, $R_{n;j}$, einzeln berechnet, s. Gleichung (6-23), und die Teststatistiken folgendermaßen formuliert:

$$D_n^1 = \sup_t \|\mathbf{R}_n^1(t_i)\|_2 \quad \text{Kolmogoroff-Smirnov-Teststatistik,} \quad (6-27)$$

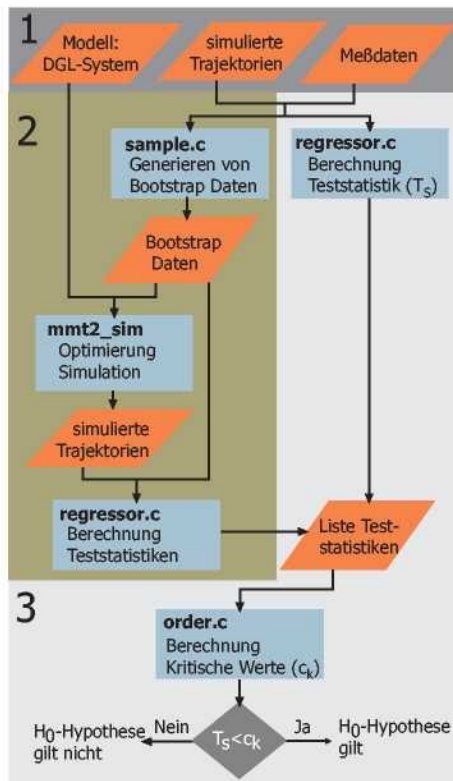
$$[W_n^1]^2 = n^{-1} \sum_{i=1}^n \|\mathbf{R}_n^1(t_i)\|_2^2 \quad \text{Cramér-von-Mises-Teststatistik,} \quad (6-28)$$

$$\text{mit } \mathbf{R}_n^1(t) = [R_{n;1}^1(t), \dots, R_{n;m}^1(t)]^T.$$

6.3.3 Implementierung

Die Arbeitsschritte für die Durchführung des im letzten Abschnitt beschriebenen Hypothesentests zur Prüfung der Regressionsgüte eines Modells lassen sich folgendermaßen zusammenfassen:

1. Anpassung des zu prüfenden Modells an den originalen Datensatz;
2. Erzeugung der Bootstrap-Datensätze ($200 \leq b \leq 500$);
3. Berechnung der K.S.- bzw. CvM-Teststatistik aus den Datensätzen;
4. Anpassung des zu prüfenden Modells an die b Wild-Bootstrap-Datensätze;
5. Berechnung der b K.S.- bzw. CvM-Teststatistiken aus dem empirischen Prozeß der Residuen für die Bootstrap-Datensätze;



```

function MAIN_PROB(Modell, SimDaten(t, Spezies),
  OrigDaten(t,Spezies),b,alpha)

  PruefGroesse←
    REGRESSOR(SimDaten(t, Spezies),OrigDaten(t, Spezies));
  for i=1 to b do
    BootDaten(t,Spezies)←
      SAMPLE(OrigDaten(t,Spezies),SimTrajektorien(t,Spezies));
    XMLDatei←CONNECT(Modell,BootDaten(t,Spezies));
    Compilieren des Modells: mmt2_sim entsteht
    MMT2_MKSIM(XMLDatei);
    StartParaVektor←
      RANDOMIZE(UntereGrenze,ObereGrenze,OrigParaVektor);
    Anpassen des Modells an die Bootstrap-Daten
    OptParaVektor←MMT2_SIM - -FIT(StartParaVektor);
    Simulation mit optimierten Parametern
    SimBoot(t,Spezies)←MMT2_SIM - -SIM(OptParaVektor);
    Liste(b,Ts)←
      REGRESSOR(SimBoot(t,Spezies),BootDaten(t,Spezies));
  end;
  KritWert←ORDER(Liste(),alpha);
  AUSGABE(KritWert,PruefGroesse);
end;

```

Abb. 6–5 Implementierung des Bootstrap-Verfahrens; links: Flußdiagramm für den vereinfachten Ablauf des Verfahrens; rechts: vereinfachter Algorithmus der Hauptroutine;

6. Bildung der Rangordnung und Berechnung der kritischen Werte;

7. Vergleich der Prüfgröße mit dem kritischen Wert $\Rightarrow$ Akzeptieren bzw. Ablehnen der H_0 -Hypothese;

In Abbildung 6–5 (links) ist ein Flußdiagramm für die von der Prozedur durchzuführenden Arbeitsschritte aufgezeichnet. In Teil 1 des Flußdiagramms sind die für das Verfahren notwendigen Eingabedateien aufgeführt: a) eine XML-Rumpfdati, die das DGL-System des zu prüfenden Modells und die optimierten Modellparameter beinhaltet; b) eine ASCII-Datei mit den Meßdaten; c) eine ASCII-Datei mit den simulierten Trajektorien des an die Meßdaten angepaßten Modells. In Teil 2 und Teil 3 des Flußdiagramms ist der Ablauf des Verfahrens dargestellt, zu dessen Realisierung verschiedene C-Routinen in ein Perl-Skript „main_prob.pl“ eingebettet wurden, s. Abb. 6–5 (rechts) für eine vereinfachte Übersicht über das Skript.

Die Arbeitsschritte in Teil 2 des Ablaufdiagramms werden für jeden Bootstrap-Datensatz einmal, d.h. insgesamt also b Mal, ausgeführt. Dabei liest die C-Routine „sample.c“ die Meßdatendatei bzw. die Datei mit den Simulationsergebnissen ein und berechnet aus ihnen die Bootstrap-Daten^{6.13}. Diese werden über das Perl-Skript mit der XML-Rumpfdati, die das DGL-System des Modells enthält, zusammengesetzt, um eine Eingabedatei für MMT2 zu erzeugen. Zur Parameteranpassung des Modells an den Bootstrap-Datensatz werden die MMT2-Routinen „mmt2_mksim“ bzw. „mmt2_sim“ mit der XML-Eingabedatei und zufällig erzeugten^{6.14} Startwerten für die Parameter aufgerufen. Im Anschluß an die Optimierung

^{6.13}Die hierfür benötigten Zufallszahlen aus einer $N(0,1)$ -Verteilung wurden unter Verwendung der g05ddc-Routine der NAG C Bibliothek^{4,7} erzeugt.

^{6.14}Die Zufallszahlen für jeden Parameter p_j wurden einer Gleichverteilung mit einer konstanten Wahrscheinlichkeitsdichte innerhalb des Bereiches: $0,5p_j \leq p_j \leq 1,5p_j$, entnommen. Zur Erzeugung der Zahlen wurde die Funktion g05dac aus der NAG C Bibliothek aufgerufen.

berechnet die C-Routine „regressor.c“ aus den simulierten Trajektorien des angepaßten Modells die jeweilige R_n^* -Funktion und aus dieser die Teststatistiken.

Die gleiche Routine wird zur Berechnung der entsprechenden Teststatistiken für das an den Originaldatensatz angepaßte Modell aufgerufen, s. Teil 3 des Flußdiagramms. Abschließend erstellt die Routine „order.c“ die Rangordnung der Teststatistiken für die b -Bootstrap-Datensätze, ermittelt die kritischen Werte $c_k(\alpha)$ für eine gegebene Irrtumswahrscheinlichkeit α und vergleicht sie mit den Prüfgrößen. Daraus resultiert die Annahme oder die Ablehnung der H_0 -Hypothese.

Die Implementation des Verfahrens wurde anhand des ersten Beispiels in (Stute et al. 1998) verifiziert, d.h. es ergaben sich bei Verwendung der hier erstellten Routinen Resultate, die den in (Stute et al. 1998) beschriebenen Ergebnissen für das univariate Regressionsmodell Nummer 1 entsprechen.

6.3.4 Ergebnisse für ein einfaches Stoffwechselmodell

Die Anwendbarkeit des Verfahrens auf Stoffwechselmodelle wurde mit einem einfachen makrokinetischen Modell untersucht (Munack 1989), s. Abb. 6–6 (oben). Ausschlaggebend für die Wahl dieses Beispielsystems war, daß die betreffenden Modellparameter als praktisch identifizierbar beschrieben wurden (Munack 1989). Eine Analyse der lokalen Sensitivitätsfunktionen bzgl. der verwendeten Parameter wies jedoch auf eine relativ geringe Sensitivität der Ausgabefunktionen bzgl. der Konstanten K_s , s. Abb. 6–6 (unten), hin, so daß sie bei allen Parameteranpassungen des Modells auf ihren Nominalwert festgesetzt wurde.

Zur Prüfung des Verfahrens wurde aus dem in Gleichung (6–29) beschriebenen Modell eine größere Zahl ($d = 1000$) an *in-silico*-Datensätzen erzeugt, indem die Ausgabefunktionen mit einem normal- oder uniform-verteilten, heteroskedastischem Rauschen $\mathbf{w}(t_i)$ überlagert wurden:

$$\begin{aligned}\tilde{\mathbf{y}}(t_i) &= \mathbf{y}(t_i) + \mathbf{w}(t_i), \\ \mathbf{w}(t_i) &\sim N(0, \boldsymbol{\sigma}^2(t_i)) && \text{(Normalverteilung),} \\ \mathbf{w}(t_i) &\sim U(-\Delta\mathbf{y}(t_i), \Delta\mathbf{y}(t_i)) && \text{(Gleichverteilung),} \\ \Delta\mathbf{y}(t_i) &= \boldsymbol{\sigma}(t_i) = f_{rel} \cdot \left[1 + \left(\frac{\mathbf{y}(t_i)}{ug} + \left(\frac{\mathbf{y}(t_i)}{ug} \right)^2 \right)^{-1} \right],\end{aligned}\tag{6–30}$$

wobei $\mathbf{y}(t_i) = [s(t_i), x(t_i)]^T$ ist, und ug auf den Wert 0,05 g/l gesetzt wurde, s. (Wiechert 1990; Takors 1997) für die Beziehung zur Berechnung der Meßunsicherheit $\Delta\mathbf{y}(t_i)$. Für diese *in-silico* Datensätze stellt das Modell aus Gleichung (6–29) definitiv das richtige Modell dar; d.h. wird es an die *in-silico*-Datensätze angepaßt, sollte es folglich auch von dem abgewandelten Bootstrap-Verfahren in einer überwiegenden Zahl der Fälle akzeptiert werden. Weiterhin gilt gemäß den Gesetzen der Statistik, daß die Zahl der fälschlich von dem Verfahren abgelehnten Modelle um so geringer sein sollte, je höher die Zahl der Datenpunkte n und je kleiner der Rauschanteil bei der Erstellung der *in-silico*-Datensätze gewählt wird.

Die Ergebnisse sind in Tabelle 6–3 zusammengefaßt. Es zeigt sich, daß die prognostizierten Trends bzgl. der Datenpunkte und der Stärke des Rauschens bis auf kleinere Abweichungen tatsächlich beobachtet werden können. Für Modelle, die an Datensätze mit einem überlagerten, normalverteilten Rauschen angepaßt wurden, zeigt das Testverfahren leicht bessere Ergebnisse als bei Datensätzen mit uniform verteiltem Rauschen. Dies kann evtl. damit begründet werden, daß die Residuen $\epsilon(t_i)$ näherungsweise mit dem Rauschanteil $\mathbf{w}(t_i)$ gleichgesetzt werden können, wenn das wahre Modell an die aus ihm

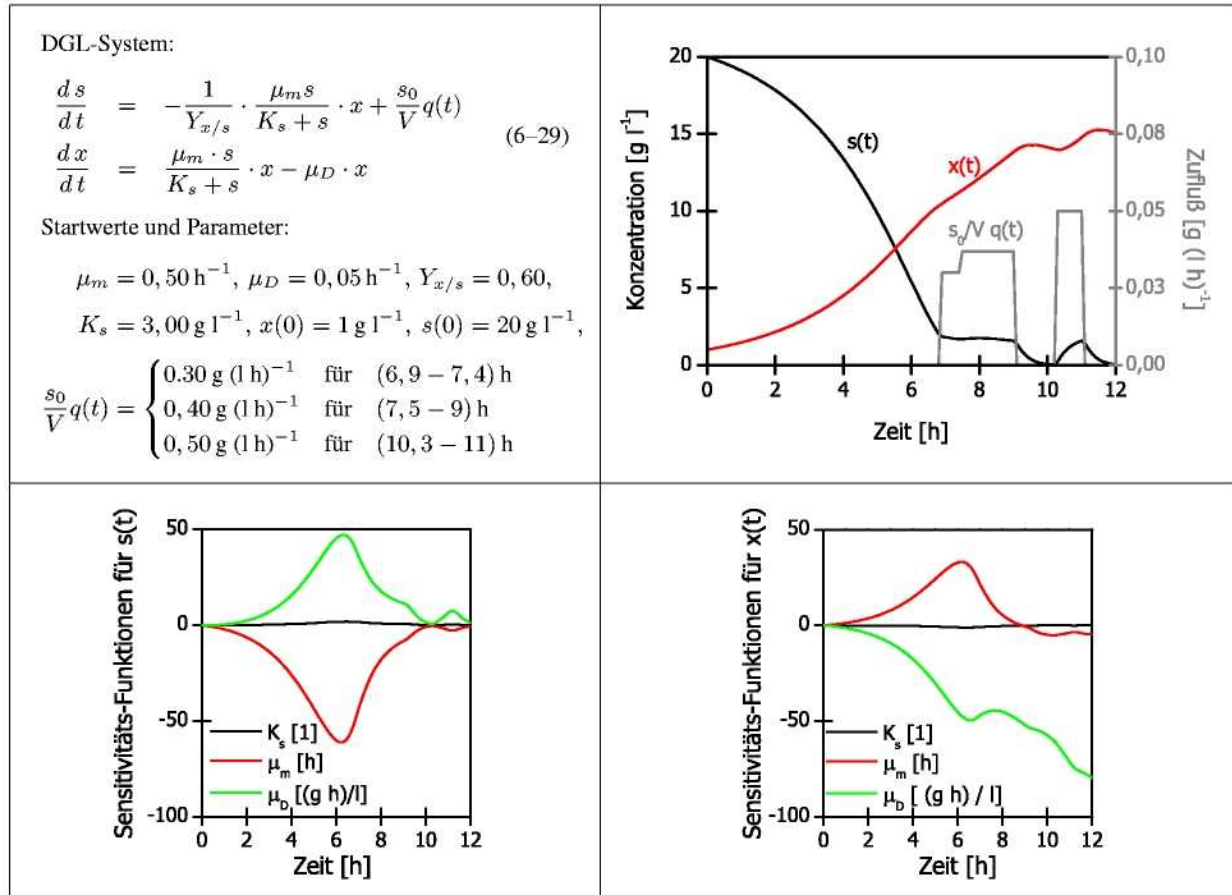


Abb. 6–6 Makrokinetisches Stoffwechselmodell (Munack 1989) zur Untersuchung der Anwendbarkeit des Anpassungstests auf Basis der Bootstrap-Datensätze; oben links: zugrundeliegendes DGL-System mit den verwendeten Modellparametern und Startwerten; oben rechts: simulierte Ausgabefunktionen; unten links: lokale Sensitivitätsfunktionen der Substratkonzentration, bzw. unten rechts: der Biomassekonzentration in Abhängigkeit von den Modellparametern

erzeugten Meßdaten angepaßt wird. Somit entspricht die Verteilungsfunktion des residualen empirischen Prozesses bei Meßdaten mit normalverteilten Streuungen unabhängig von der gewählten Zahl an Datenpunkten n bereits einer Normalverteilung, während bei gleichverteiltem Rauschen erst ab einem genügend hohen Wert für n ein Grenzprozeß mit Normalverteilung angenähert wird.

Insgesamt werden jedoch nie mehr als ca. 20% der Modelle für eine sehr geringe Zahl an Datenpunkten abgelehnt, und auch die in Tab. 6–3 aufgeführten P-Werte weisen auf eine hohe Akzeptanz des wahren Modells durch den Anpassungstest hin. Der P-Wert gibt die Wahrscheinlichkeit an, eine mindestens so große Prüfgröße wie die aus den Daten berechnete zu erhalten, wenn die Nullhypothese gilt. Im Gegensatz zu dem vor Testbeginn festzulegenden Wert für die Irrtumswahrscheinlichkeit α entspricht er also der empirisch ermittelten Wahrscheinlichkeit, mit der die Nullhypothese gerade noch widerlegt werden kann, s. (Sachs 2004, S. 188). Damit ergibt sich der P-Wert für den Hypothesentest des Bootstrap-Verfahrens als der über d Datensätze arithmetisch gemittelte Bruchteil der aus den Bootstrap-Daten berechneten Teststatistiken, die größer als die Prüfgröße sind:

$$\text{P-Wert} = \frac{1}{d} \sum_{o=1}^d b^{-1} \# \{D_n^{*,q} > D_n\}_o. \quad (6-31)$$

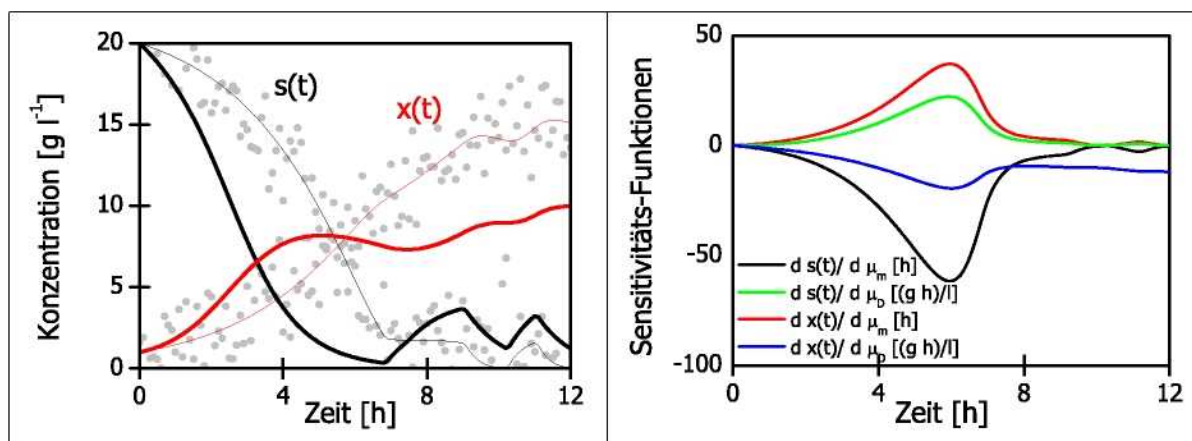
Es zeigt sich in Tab 6–3, daß der P-Wert immer größer als 0,45, d.h. sehr viel höher als das vorher festgelegte Signifikanzniveau von $\alpha = 0,05$ ist. Die Nullhypothese wird also deutlich bestätigt.

Tab. 6–3 P-Wert und Prozentsatz der abgelehnten (wahren) Modelle aus Gleichung (6–29) in Abhängigkeit von der Zahl der Datenpunkte und der Größe des Rauschanteils (1000 Datensätze, 500 Bootstrap-Datensätze)

		Gleichverteilung des Rauschens					
		$f_{\text{rel}}=1,5$		$f_{\text{rel}}=2,5$		$f_{\text{rel}}=5,5$	
n		K.S.	CvM	K.S.	CvM	CvM	K.S.
16	Prozent	17,8 %	17,8 %	19,3 %	19,3 %	17,6 %	17,6 %
	P-Wert	0,459	0,459	0,467	0,464	0,468	0,468
61	Prozent	15,9 %	15,9 %	17,6 %	17,6 %	16,8 %	16,8 %
	P-Wert	0,487	0,495	0,482	0,491	0,484	0,486
121	Prozent	10,4 %	10,4 %	12,8 %	12,8 %	14,2 %	14,2 %
	P-Wert	0,490	0,493	0,498	0,498	0,485	0,498
241	Prozent	8,7 %	8,7 %	9,9 %	9,9 %	9,9 %	9,9 %
	P-Wert	0,495	0,502	0,492	0,493	0,493	0,497

		Normalverteilung des Rauschens					
		$f_{\text{rel}}=1,5$		$f_{\text{rel}}=2,5$		$f_{\text{rel}}=5,5$	
n		K.S.	CvM	K.S.	CvM	CvM	K.S.
16	Prozent	4,2%	4,2%	5,5%	5,5%	4,1%	4,1%
	P-Wert	0,534	0,537	0,503	0,513	0,535	0,533
61	Prozent	5,8%	5,8%	4,9%	4,9%	3,8%	3,8%
	P-Wert	0,532	0,544	0,521	0,527	0,538	0,550
121	Prozent	4,9%	4,9%	4,7%	4,7%	6,3%	6,3%
	P-Wert	0,539	0,536	0,530	0,540	0,519	0,523
241	Prozent	3,8%	3,8%	5,4%	5,4%	4,1%	4,1%
	P-Wert	0,544	0,544	0,530	0,545	0,532	0,503

Damit ist die „schwache“ Aussage des auf den Bootstrap-Daten basierenden Testverfahrens in einer Vielzahl von Fällen zurechtgetroffen worden, d.h. die Nullhypothese wird beibehalten, wenn sie richtig ist, und infolgedessen das richtige Modell mit einer hohen Frequenz akzeptiert.

**Abb. 6–7** Abgewandelter Stoffwechselmodell aus Gleichung (6–32); links: simulierte Trajektorien, dicke Linien: abgewandelter Modell, dünne Linien: ursprüngliches Modell, graue Punkte: über Gleichung (6–30) erzeugter Datensatz; rechts: Sensitivitätsfunktionen bzgl. der Modellparameter

Eine Nullhypothese abzulehnen, gilt allerdings als stärkere Aussage, besonders, wenn sie deutlich falsch ist, s. (Sachs 2004). Um das Bootstrap-Testverfahren dahingehend zu prüfen, wurde nur die Beziehung

Tab. 6–4 Vergleich verschiedener Kriterien zur Beurteilung der Anpassungsgüte; gemittelte Ergebnisse aus 500 Anpassungen des wahren Stoffwechselmodells, s. Gleichung (6–29), und des abgewandelten Modells, s. Gleichung (6–32), an über Gleichung (6–30) erzeugte Datensätze (normalverteilte Streuungen der Datenpunkte, $f_{rel}=2,5$, $b=300$); s. die Diskussion auf den Seiten 112 bis 114 zur Bewertung der Ergebnisse für die Fehlerquadratsumme, die Likelihood-Funktion und das Akaike-Kriterium

Modell aus Gleichung (6–29) (wahres Modell)				
Bootstrap-Testverfahren (K.S.-Statistik); Dimension [g/l]				
Prüfwert: 2,059	Krit. Wert: 3,416	Prozent: 4,2%	P-Wert:0,521	F^a [(g/l) ²] 4,599 $\ln(\mathcal{L})^b$ -441,519 AIK^c 887,038
Bootstrap-Testverfahren (CvM-Statistik); Dimension [g/l]				
Prüfwert: 1,611	Krit. Wert: 4,635	Prozent: 4,2%	P-Wert:0,529	
Modell aus Gleichung (6–32) („falsches Modell“)				
Bootstrap-Testverfahren (K.S.-Statistik); Dimension [g/l]				
Prüfwert: 35,803	Krit. Wert: 18,013	Prozent: 98,2%	P-Wert:0,01	F^a [(g/l) ²] 42,294 $\ln(\mathcal{L})^b$ -599,070 AIK^c 1202,139
Bootstrap-Testverfahren (CvM-Statistik); Dimension [g/l]				
Prüfwert: 725,559	Krit. Wert: 110,227	Prozent: 98,2%	P-Wert:0,01	

^a F : Fehlerquadratsumme
^b $\ln \mathcal{L}$: Wert der Log-Likelihood-Funktion
^cAIK: Akaike-Kriterium

für $x(t)$ aus dem DGL-System in Gleichung (6–29) folgendermaßen abgewandelt:

$$\frac{dx}{dt} = \frac{k}{Y_{x/s}} \cdot s \cdot x - \mu_D, \quad (6-32)$$

und die übrigen Gleichungen sowie das Profil des Substratzuflusses beibehalten. Selbst nach der Bestimmung der optimalen Parameterwerte ($k = 0,03 \text{ l / (g h)}$, $\mu_D = 0,184 \text{ h}^{-1}$) kann dieses Modell die über Gleichung (6–30) aus dem ursprünglichen Modell erzeugten *in-silico*-Datensätze nur noch tendenziell beschreiben, s. Abb. 6–7. Es sollte daher auch mit einer hohen Frequenz durch das Bootstrap-Testverfahren abgelehnt werden. In Tab. 6–4 sind die Ergebnisse für 500 Datensätze, die aus 121 verrauschten Datenpunkten bestehen ($\mathbf{w}(t_i) \sim N(0, \sigma^2(t_i))$, $f_{rel}=2,5$) im Vergleich mit den entsprechenden Ergebnissen des Originalmodells dargestellt. Unabhängig von der gewählten Teststatistik wird das Modell in 491 von 500 Fällen verworfen, die Nullhypothese also deutlich abgelehnt. Die P-Werte sind entsprechend klein, d.h. bereits auf einem 1%-Signifikanzniveau hätte die Nullhypothese widerlegt werden können. Bei dem Vergleich der gemittelten Prüf- und kritischen Werte wird ersichtlich, daß die Zahlen für das angepaßte, abgewandelte Modell insgesamt höher sind als für das richtige Modell. Dies resultiert aus der Tatsache, daß die Teststatistiken Funktionen der Verteilung des residualen empirischen Prozeß sind und somit direkt von der Varianz der Residuen abhängen. Der über die CvM-Teststatistik berechnete durchschnittliche kritische Wert für das angepaßte richtige Stoffwechselmodell ist 2,9-mal höher als der Prüfwert, während der aus der K.S.-Teststatistik berechnete Wert nur 1,7-mal höher ist. Für das abgewandelte Stoffwechselmodell kehrt sich das Verhältnis um, d.h. der über die CvM-Statistik ermittelte Quotient aus Prüf- und kritischem Wert ist deutlich höher als das korrespondierende Resultat aus der K.S.-Teststatistik. Unter Verwendung der CvM-Statistik wird also das richtige Modell „eindeutiger“ akzeptiert bzw. das weniger gut angepaßte Modell „deutlicher“ zurückgewiesen. Dies könnte auf leicht bessere Ergebnisse eines auf der Cramér-von-Mises Teststatistik beruhenden Anpassungstest hinweisen, was auch schon für die Regressionsmodelle in (Stute et al. 1998) beobachtet wurde.

In der Tabelle sind ebenfalls die in der Einleitung des Abschnitts erwähnten Kriterien, d.h. die Mittelwerte der Fehlerquadratsumme, der Log-Likelihood-Funktion und des AIK, für die beiden Stoffwechselmodelle enthalten.

Die Likelihood-Funktion eines multivariaten Regressionsmodells gibt die bedingte Wahrscheinlichkeit

an, die Residuen aller abhängigen Variablen $\epsilon_i = [(\tilde{y}_1 - y_1), \dots, (\tilde{y}_m - y_m)]_i^T$ für verschiedene Ausprägungen der Regressoren $[x_1, \dots, x_m]_i^T, i = 1, \dots, n$ zu beobachten, wenn der aus der Anpassung des Modells gefundene Parametervektor $\hat{\mathbf{p}}$ der wahre Vektor ist, s. z.B. (Johnson und Wichern 2002, S. 168) für die multivariate Likelihood-Funktion. Sie kann entsprechend für die Lösungen eines gewöhnlichen DGL-Systems erster Ordnung aufgestellt werden:

$$\{\mathcal{L}(\epsilon(t_i) | \hat{\mathbf{p}})\}_{i=1, \dots, n} = \prod_{i=1}^n \frac{1}{(2\pi)^{m/2} |\Sigma(t_i)|^{1/2}} e^{-(\tilde{\mathbf{y}}(t_i) - \mathbf{y}(t_i))^T \Sigma(t_i)^{-1} (\tilde{\mathbf{y}}(t_i) - \mathbf{y}(t_i)) / 2}, \quad (6-33)$$

wobei die Annahme gilt, daß die zu verschiedenen Zeitpunkten ermittelten Residuen normalverteilt und wechselseitig unabhängig sind. $\Sigma(t_i)$ in Gleichung (6-33) steht für die Kovarianzmatrix der Residuen:

$$\begin{aligned} \Sigma(t_i) &= \mathbb{E} [\tilde{\mathbf{y}}(t_i) - \mathbf{y}(t_i)] [\tilde{\mathbf{y}}(t_i) - \mathbf{y}(t_i)]^T \\ &= \mathbb{E} \left(\begin{bmatrix} \tilde{y}_1(t_i) - y_1(t_i) \\ \vdots \\ \tilde{y}_m(t_i) - y_m(t_i) \end{bmatrix} [\tilde{y}_1(t_i) - y_1(t_i), \dots, \tilde{y}_m(t_i) - y_m(t_i)] \right). \end{aligned} \quad (6-34)$$

Wird der Korrelationskoeffizient: $\rho_{12}(t_i) = \sigma_{12}(t_i) / (\sqrt{\sigma_{11}(t_i)} \cdot \sqrt{\sigma_{22}(t_i)})$, eingeführt, lautet Gleichung (6-33) für das vorliegende bivariate System wie folgt:

$$\begin{aligned} \{\mathcal{L}(\epsilon_1(t_i), \epsilon_2(t_i) | \hat{\mathbf{p}})\}_{i=1, \dots, n} &= \prod_{i=1}^n \frac{1}{2\pi \sqrt{\sigma_{11}(t_i) \sigma_{22}(t_i) (1 - \rho_{12}(t_i)^2)}} \times \\ &\exp \left\{ -\frac{1}{2(1 - \rho_{12}(t_i)^2)} \left[\left(\frac{\tilde{y}_1(t_i) - y_1(t_i)}{\sqrt{\sigma_{11}(t_i)}} \right)^2 + \left(\frac{\tilde{y}_2(t_i) - y_2(t_i)}{\sqrt{\sigma_{22}(t_i)}} \right)^2 \right. \right. \\ &\quad \left. \left. - 2\rho_{12}(t_i) \left(\frac{\tilde{y}_1(t_i) - y_1(t_i)}{\sqrt{\sigma_{11}(t_i)}} \right) \left(\frac{\tilde{y}_2(t_i) - y_2(t_i)}{\sqrt{\sigma_{22}(t_i)}} \right) \right] \right\}. \end{aligned} \quad (6-35)$$

Die zur Berechnung von Gleichung (6-35) benötigten Schätzer für die residualen Varianzen und Kovarianzen ergeben sich definitionsgemäß zu:

$$\begin{aligned} \hat{\sigma}_{11}(t_i)^2 &= \frac{1}{d-q} \sum_{o=1}^d (\tilde{y}_{1,o}(t_i) - y_{1,o}(t_i))^2, \\ \hat{\sigma}_{22}(t_i)^2 &= \frac{1}{d-q} \sum_{o=1}^d (\tilde{y}_{2,o}(t_i) - y_{2,o}(t_i))^2, \\ \hat{\sigma}_{12}(t_i)^2 &= \frac{1}{d-q} \sum_{o=1}^d (\tilde{y}_{1,o}(t_i) - y_{1,o}(t_i)) \cdot (\tilde{y}_{2,o}(t_i) - y_{2,o}(t_i)), \end{aligned} \quad (6-36)$$

d.h. die quadrierten Residuen zum Zeitpunkt t_i werden für die Datensätze $o = 1, \dots, d$ gemittelt, wobei die Zahl der Freiheitsgrade der Zahl der unabhängigen Messungen zum Zeitpunkt t_i erniedrigt um die Zahl der angepaßten Parameter q entspricht. Als Kriterium zur Beurteilung der Anpassungsgüte wird

in der Regel der Wert der Log-Likelihood-Funktion angegeben, unter anderem weil die Likelihood-Funktion über mehrere 100er-Potenzen variieren kann:

$$\begin{aligned} \ln(\{\mathcal{L}(\epsilon_1(t_i), \epsilon_2(t_i) | \hat{\mathbf{p}})\}_{i=1, \dots, n}) &= -\ln \frac{n}{2} - \frac{m \cdot n}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma(t_i)| \\ &\quad - \sum_{i=1}^n (\tilde{\mathbf{y}}(t_i) - \mathbf{y}(t_i))^T \Sigma(t_i)^{-1} (\tilde{\mathbf{y}}(t_i) - \mathbf{y}(t_i)). \end{aligned} \quad (6-37)$$

Je größer die über Gleichung (6-35) oder (6-37) berechnete Zahl ist, desto höher wird die Anpassungsgüte des Modells bewertet. Umgekehrt können optimale Parameterwerte des Modells über die Maximierung der (Log-)Likelihood-Funktion gefunden werden^{6,15}. Der in Tab. 6-4 aufgeführte Wert der Log-Likelihood-Funktion ist bereits für das richtige Stoffwechselmodell stark negativ. Dies läßt sich vorwiegend darauf zurückführen, daß bei vernachlässigbaren Kovarianzen die residualen Varianzen $\sigma_{11}(t_i) \approx \sigma_{22}(t_i)$, $i = 1, \dots, n$ einen Wert von ca. 2 (g/l)² besitzen, s. Gleichung (6-30), und daher der Vorfaktor des Exponenten zu jedem Zeitpunkt kleiner als eins ist. Die Log-Likelihood-Funktion des abgewandelten Modells ist aufgrund der mit ihm erreichten, geringeren Wiedergabequalität der Daten jedoch deutlich kleiner, so daß die Modelle auch über die Log-Likelihood-Funktion eindeutig klassifiziert werden können.

Wie von Akaike (Akaike 1973) festgestellt wurde, stellt die Likelihood-Funktion einen nach oben hin verzerrten Schätzer eines Modellselektionskriteriums auf Basis der Anpassungsgüte dar. Weiterhin wurde in derselben Publikation gezeigt, daß die Verzerrung in der Regel der Zahl der Modellparameter entspricht, s. (Akaike 1973; Burnham und Anderson 1998). Unter Berücksichtigung eines aus „historischen Gründen“ gewählten Faktors von -2 wurde somit das Akaike-Informationen-Kriterium, AIK, vorgeschlagen:

$$\text{AIK} = -2 \ln(\mathcal{L}(\epsilon | \hat{\mathbf{p}})) + 2p, \quad (6-38)$$

das zur Modelldiskriminierung eingesetzt werden soll. Unter Modelldiskriminierung wird dabei verstanden, daß aus einer Modellfamilie, die zur Beschreibung eines Datensatzes entwickelt wurde, das Modell mit dem geringsten AIK als Bestes selektiert werden sollte, bzw. eine Rangordnung der Modelle entsprechend der AIK-Differenzen: $\Delta_z = \text{AIK}_z - \min \text{AIK}$, erstellt werden kann, wobei z der Index für die verschiedenen Modelle ist.

Da für die gewählten Beispielsysteme die Log-Likelihood-Funktionen sehr klein sind und nur zwei Parameter angepaßt wurden, unterscheiden sich die berechneten AIK in ihrer Aussage nicht deutlich von den $\ln \mathcal{L}(\epsilon | \hat{\mathbf{p}})$ -Werten. Es ist allerdings denkbar, daß besonders mit zunehmender Parameterzahl das AIK der Likelihood-Funktion zur Bewertung der Anpassungsgüte vorzuziehen ist. Es als alleiniges Kriterium zur Modellselektion zu verwenden, erscheint hier jedoch unangebracht, da andere Aspekte zur Beurteilung der Qualität eines Modells, wie z.B. die Parametersensitivitäten, völlig außer Acht gelassen werden.

6.3.5 Zusammenfassung

Die Anpassungsgüte zweier einfacher Stoffwechselmodelle wurde anhand von vier verschiedenen Kriterien (Fehlerquadratsumme, Log-Likelihood-Funktion, Akaike-Kriterium und Anpassungstest auf Basis

^{6,15}In der Praxis wird die Varianz-Kovarianzmatrix häufig durch die Meßfehler-Kovarianzmatrix, $\mathbf{C}_i$, s. Gleichung (6-17), ersetzt. In diesem Fall sind die ersten drei Terme von Gleichung (6-37) konstant und nur der letzte Term muß maximiert werden. Dieser entspricht dem Gütefunktional der verallgemeinerten Regression, s. Gleichung (6-16).

der Bootstrap-Datensätze) bewertet. Es wurde gezeigt, daß alle Kriterien generell geeignet zur Prüfung der Anpassungsgüte von Stoffwechselmodellen sind.

Als bestes Kriterium zur Modellselektion wird das Testverfahren auf Basis der Bootstrap-Datensätze bewertet, da eine eindeutige Entscheidung gefällt wird, ob ein Modell akzeptiert oder verworfen werden sollte. Zusätzlich kann durch einen Vergleich der Größe des Prüf- und des kritischen Wertes bestimmt werden, wie knapp die Entscheidung ausgefallen ist. Vor Anwendung des Verfahrens sollte allerdings sichergestellt werden, daß die Modellparameter identifizierbar sind, da ansonsten das Auffinden von lokalen Optima bei der Anpassung des Modells an die Bootstrap-Datensätze die Bestimmung der kritischen Werte verfälschen könnte.

Die anderen drei Kriterien haben als Vorteil, daß sie skalare Zahlen darstellen, die ohne größeren Rechenaufwand ermittelt werden können. Sie eignen sich somit auch als zu optimierende Zielfunktionen bei der Parameteranpassung. Allerdings muß sowohl für die Berechnung der Likelihood-Funktion als auch des Akaike-Information-Kriteriums die Verteilungsfunktion der Residuen vorher bekannt sein. Ebenfalls zeigte sich bereits bei den sehr einfachen Modellen, daß die Kriterien sensitiv auf Änderungen der Größe der Varianzen reagieren. Daher sollte vor der Berechnung der Gütefunktionale der Wert der residualen Varianzen und Kovarianzen durch experimentelle Mehrfachbestimmungen abgesichert werden.

7 Zusammenfassung

Ziel der vorliegenden Arbeit war es, einen Modellierungs- und Validierungsprozeß für die Auswertung von Datensätzen aus Substrat-Puls-Experimenten zu entwickeln. Dafür wurden die Modellierungswerkzeuge MMT und MMT2 erstmalig zur Erstellung, numerischen Integration und Parameteranpassung der gewöhnlichen Differentialgleichungssysteme erster Art verwendet. Die Modellierung bezog sich auf Datensätze aus Substrat-Puls-Experimenten mit einem *E.coli* K12 Wildstamm (Experiment WT1 (Buchholz 2001)) bzw. zwei genetisch modifizierten Stämmen, d.h. *E.coli*-L-Phenylalanin-Produktionsstämmen (Experimente MU1 und MU2 (Oldiges 2004)).

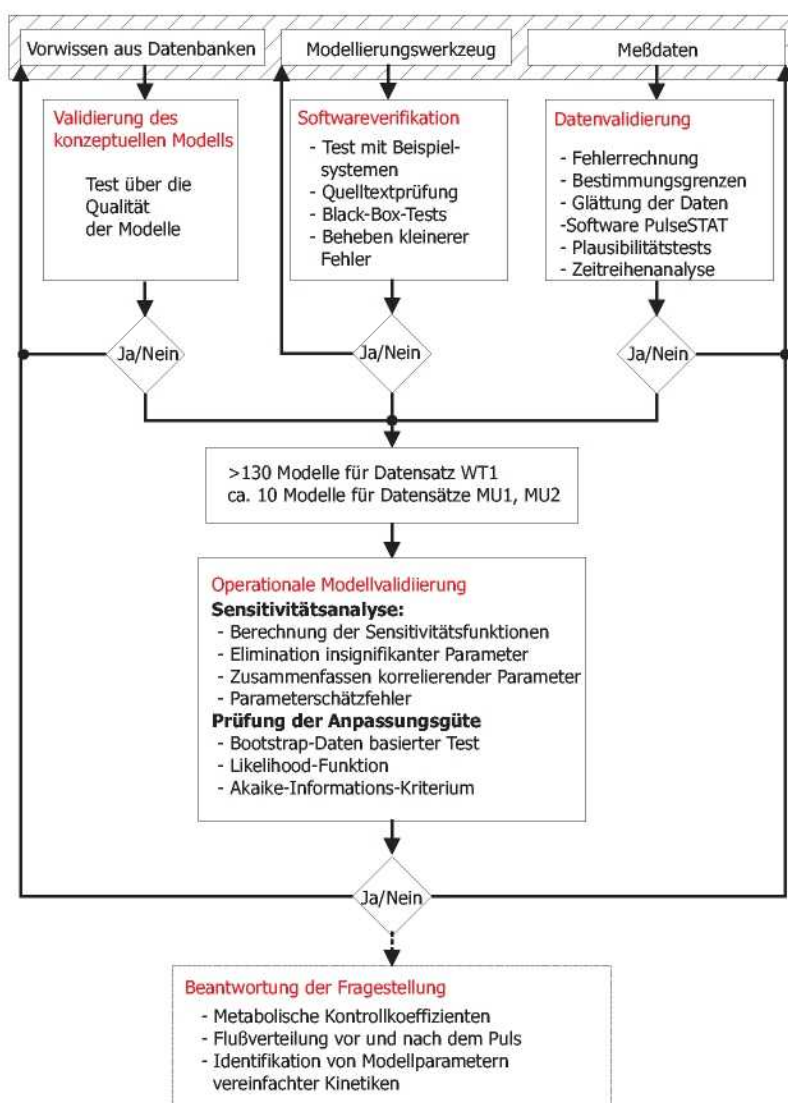


Abb. 7–1 Hauptergebnisse der vorliegenden Dissertation, entwickelte Methoden zur Beantwortung der in der Zielstellung formulierten Fragen, s. Abb. 1–2

Für die Entwicklung des Prozesses wurden bewährte Konzepte und Methoden der Modellbildung und Validierung aus dem Bereich der Modellierung und Simulation angewendet, s. Abb. 7–1. Der hier vorgeschlagene, konkrete Prozeß orientiert sich dabei an dem Modellierungszyklus aus (Sargent 1982), in dessen Mittelpunkt die Meßdaten stehen.

Eine notwendige Grundvoraussetzung für die Modellierung ist die **Validität der Daten**. Aus diesem

Grund wurden zahlreiche Methoden ausgearbeitet, die sowohl zur Prüfung der Arbeitsmethoden zur Datenerfassung als auch zur Untersuchung der Datenqualität einsetzbar sind:

- eine **Größtfehlerabschätzung** bildete die Grundlage zur Bestimmung der Meßunsicherheiten;
- die **Nachweis-, Erfassungs und Bestimmungsgrenzen**, welche maßgeblich für die Beurteilung von Arbeitsmethoden aus der quantitativen Chemischen Analytik sind, wurden für die hier vorliegenden LC-MS-Messungen auf die gängige Weise gemäß (DIN 32 645 1994) berechnet. Ziel war, einheitliche Kriterien zu erhalten, auf deren Basis verschiedene Datensätze bzw. verschiedene Meßreihen eines Datensatzes miteinander verglichen werden können;
- zur vereinfachten Auswertung der Meßsignale, zur Durchführung der Fehlerrechnung und verschiedener statistischer Tests, sowie zur Berechnung der Nachweis-, Erfassungs- und Bestimmungsgrenzen, wurde die **VBA-Anwendung PulseSTAT** programmiert. Über ihre Verwendung kann die zur Datenauswertung benötigte Zeit und die Zahl der dabei auftretenden Rechenfehler deutlich verringert werden;
- die Datensätze wurden **Plausibilitätstests** unterzogen, indem die im Fließgleichgewichtszustand bestimmten Metabolitkonzentrationen mit veröffentlichten Ergebnissen aus Substrat-Puls-Experimenten verglichen wurden. Dabei zeigte sich eine starke Abhängigkeit der experimentellen Ergebnisse von der jeweils gewählten Arbeitsmethode zur Probennahme, Zellextraktion und Metabolitquantifizierung;
- für den **Konsistenztest** der Datensätze wurde eine in (Arkin und Ross 1995; Samoilov et al. 2001) abgeleitete Variante der **Zeitreihenanalyse** herangezogen, die auf der Basis der Ähnlichkeit zeitabhängiger Meßreihen von chemischen Spezies das zugrundeliegende Reaktionsnetzwerk erschließt. Die Methode wurde implementiert, und anhand von *in-silico*-Datensätzen gezeigt, daß das Verfahren zur Detektion von dem System nicht per-se innewohnendem Rauschen sowie zur Bestimmung von inkohärenten Meßreihen geeignet ist. Bei der Anwendung der Methode auf die vorliegenden Meßdaten wurde festgestellt, daß beide Störeinflüsse alle drei Datensätze in mehr oder minder großem Ausmaß beeinträchtigen;
- verschiedene **Datenglättungsmethoden** wurden auf ihre Fähigkeit zur Elimination des Rauschens miteinander verglichen. Die **FFT-Filterung** und die **Datenapproximation über gewichtete kubische B-Splines** erwiesen sich als geeignetste Methoden und wurden so implementiert (Froemel 2003), daß künftig eine einfache, automatisierte Datenaufbereitung möglich ist. Die geglätteten Datensätze gingen in die Modellierung ein, da auf diese Weise Konvergenzschwierigkeiten des Optimierungsalgorithmus bei der Parameteranpassung der Modelle vermieden werden konnten;

Neben der Validität der Daten ist vor Beginn der Modellierung immer die Richtigkeit der zur Modellierung verwendeten Software abzusichern (Page 1991). Daher wurden zur **Verifikation von MMT und MMT2** Methoden, wie die Quelltextprüfung oder Black-Box-Tests mit Beispielmolekülen, angewendet, wobei kleinere Fehler aus MMT behoben und allgemein die Gültigkeit der Simulationsergebnisse sichergestellt werden konnte.

Über **130 Modelle wurden für den Datensatz WT1** unter Einbeziehung von Angaben aus Online-Datenbanken, wie Kegg (Kanehisa und Goto 2000), Brenda (Schomburg et al. 2002), ExPASy (Appel et al. 1994), EcoCyc (Karp et al. 2000), etc., konstruiert. Nach der Parameteranpassung der Modelle wurde anhand der Wiedergabequalität der Meßdaten im Sinne der **konzeptuellen Modellvalidierung** entschieden, ob eine geeignete Wahl der zugrundeliegenden Hypothesen und Vereinfachungen getroffen wurde. Infolgedessen konnten insgesamt 12 Modelle mit einer Dimension von 10-12 DGL und 90-130 Parametern selektiert werden, die einen Großteil der experimentellen Konzentrationsverläufe adäquat

beschrieben. Auf die gleiche Weise wurde eine **Familie aus ca. 10 Modellen zur simultanen Beschreibung der Datensätze MU1 und MU2** erstellt, von denen eines in der Lage war, die Konzentrationen der AAA-Metabolite größenordnungsmäßig wiederzugeben. Im Zuge des iterativen Prozesses zur Modellerstellung konnte gezeigt werden, daß es über das Kriterium der Anpassungsgüte zwar gelingt, inadäquate Modellstrukturen auszuschließen, aber mehrere, diesbezüglich gleichberechtigte Modelle übrigbleiben.

Daher wurden **Methoden der operationalen Modellvalidierung** mit dem Ziel herangezogen, aus den 12 Modellen für das Experiment WT1 wenige, vereinfachte und damit praktisch identifizierbare Modelle zu selektieren, die sich zur Ableitung von stimmigen Aussagen im Sinne des Metabolic Engineering eignen:

- da für die operationale Modellvalidierung keine weiteren, unabhängigen Datensätze vorhanden waren, wurde eine explorative Methode, die **Sensitivitätsanalyse**, eingesetzt, die allgemein das Mittel der Wahl für die Detektion von vernachlässigbaren Modellvariablen und zur Prüfung der Modellstruktur darstellt:
 - die **Berechnung der Sensitivitätsfunktionen** wurde erstmalig in MMT implementiert;
 - aus den Sensitivitätsfunktionen wurde das **Kriterium der durchschnittlichen Sensitivität** abgeleitet (Turányi 1990), auf dessen Basis ca. die **Hälfte der Parameter eines Modells eliminiert** werden konnten, ohne daß sich die Anpassungsgüte verschlechterte. Die Reduktion ermöglichte die **Detektion von Modellstrukturproblemen** im Bereich des PPP und bzw. oder der Glykolyse, welche sich über experimentelle Untersuchungen (Oldiges 2004) auf Inkonsistenzen im Datensatz WT1 zurückführen ließen. Deshalb wurden die folgenden theoretischen Untersuchungen nur noch beispielhaft anhand eines bestehenden Modells durchgeführt, um die weitere Vorgehensweise zur operationalen Modellvalidierung aufzuzeigen;
 - neben den insignifikanten Parametern verhindern **Korrelationen zwischen Parametern** die praktische Identifizierbarkeit des Modells. Sie können über die im Rahmen dieser Arbeit implementierte Methode zur **Hauptkomponentenanalyse der Kosensitivitätsmatrix** detektiert werden (Vajda et al. 1985). Infolgedessen wurde **ein bereits reduziertes Modell über das Zusammenfassen von weiteren 50% der Parameter stark vereinfacht**. Daraus läßt sich schließen, daß generell auf der Basis eines einzelnen Datensatzes aus Substrat-Puls-Experimenten höchstens ca. drei-vier Parameter pro gemessenem und im Modell berücksichtigtem Metaboliten praktisch identifiziert werden können;
 - um die **Parameterschätzfehler** als wichtige Kennzahlen für die Güte eines Modells zu ermitteln, wurde die Berechnung der **Fisher-Informationsmatrix** (Ljung und Glad 1994a) aus den Sensitivitätsfunktionen implementiert, die sich in weiterführenden Arbeiten auch für die experimentelle Versuchsplanung einsetzen läßt;
- ergänzend zur Sensitivitätsanalyse wurde anhand von *in-silico*-Datensätzen geprüft, ob sich andere **Kriterien zur Beurteilung der Anpassungsgüte**, wie z.B. die **Likelihood-Funktion** (Fisher 1925) oder das für Regressionsmodelle bekannte **Akaike-Informations-Kriterium** (Akaike 1973), besser als die Fehlerquadratsumme zur Unterscheidung der Modelle eignen. Dabei erwies sich ein auf **Bootstrap-Datensätzen basierender statistischer Entscheidungstest** (Stute 1997; Stute et al. 1998) als eine für die Situation der Substrat-Puls-Experimente zugeschnittene und empfindliche Methode;

Insgesamt wurden im Rahmen dieser Arbeit über **150 DGL-Modelle zur Beschreibung des mikrobiellen Stoffwechselnetzwerkes** generiert. Weiterhin wurden **acht verschiedene Methoden zur Daten-**

und Modellvalidierung in knapp 20 C-Routinen, Perl-Skripten sowie einer VBA-Anwendung implementiert, und ihre Verwendbarkeit anhand der vorliegenden Daten bzw. entwickelten Modelle demonstriert. Auf der Grundlage dieser Arbeiten ist es somit möglich, schnell und effizient **valide Modelle** zu entwickeln, die:

- die Parameteridentifizierung von vereinfachten Enzymkinetiken, wie z.B. Linlog-Kinetiken, zulassen;
- die Berechnung von metabolischen Kontrollkoeffizienten ermöglichen;
- die Berechnung der Flußverteilung unter stationären und instationären Bedingungen im Sinne des Metabolic Engineering erlauben;

Weiterhin haben die Ergebnisse dieser Arbeit gezeigt, daß künftig eine drei- bis viermalige Reproduktion der Substrat-Puls-Experimente, insbesondere zur Reduktion der Meßunsicherheiten, unumgänglich ist, bevor die Datensätze der Modellierung zugrundegelegt werden sollten.

Die innerhalb der Dissertation erzielten Resultate wurden im Rahmen von **sechs Publikationen** in Zeitschriften und referierten Tagungsbänden veröffentlicht. Des weiteren fanden sie Eingang in **acht Konferenzbeiträge**.

Literaturverzeichnis

- Aarts, E. H. L.; Lenstra, J. K. (1997):** Local Search in Combinatorial Optimization. In: *Wiley-Interscience Series in Discrete Mathematics and Optimization*, Kap. 1, Wiley, Chichester, New York.
- Agius, L.; Sheratt, H. S. A.(Hrsg.) (1997):** *Channeling in Intermediary Metabolism*. Portland Press Ltd., London.
- Agrawal, A. (1999):** New institute to study systems biology. *Nature Biotechnology* **17**(8), 743–744.
- Aiken, R. C. (1985):** *Stiff computation*. Oxford University Press, New York, Oxford.
- Akaike, H. (1973):** Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N.; Csaki, F. (Hrsg.): *Second International Symposium on Information Theory*, 267–281, Budapest, Akademiai Kiado.
- Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. (2002):** *Molecular Biology of the Cell*. Garland Science, New York, USA, 4 Aufl.
- Alliance for Cellular Signaling (USA) (2004):** <http://www.AfCS.org>. Zugriffsdatum: 03/2004.
- Alm, E.; Arkin, A. P. (2003):** Biological networks. *Current Opinion in Structural Biology* **13**(2), 193–202.
- Alt, W. (2002):** *Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen..* Vieweg Verlag, Wiesbaden.
- Apostolakis, J.; Plückthun, A.; Caffisch, A. (1998):** Docking Small Ligands in Flexible Binding Sites. *Journal of Computational Chemistry* **19**(1), 21–37.
- Appel, R. D.; Bairoch, A.; Hochstrasser, D. F. (1994):** A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends in Biochemical Sciences* **19**(6), 258–260. <http://www.expasy.ch>.
- Arkin, A.; Ross, J. (1994):** Computational functions in biochemical reaction networks. *Biophysical Journal* **67**(2), 560–578.
- Arkin, A.; Ross, J. (1995):** Statistical Construction of Chemical Reaction Mechanisms from Measured Time-Series. *Journal of Physical Chemistry* **99**(3), 970–979.
- Arkin, A.; Shen, P.; Ross, J. (1997):** A Test Case of Correlation Metric Construction of a Reaction Pathway from Measurements. *Science* **277**(5330), 1275–1279.
- Arkin, A. P. (2001a):** Synthetic cell biology. *Current Opinion in Biotechnology* **12**(6), 638–644.
- Arkin, A.P. (2001b):** SIMULAC and DEDUCE software. <http://gobi.lbl.gov/aparkin/stuff/software.html>. Zugriffsdatum: 07/2001.
- Armijo, L. (1966):** Minimizing functions having Lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics* **16**, 1–3.
- Asratian, A. S.; Denley, T. M. J.; Haggkvist, R. (1998):** *Bipartite Graphs and their Applications..* Bd.131 d. Reihe *Cambridge Tracts in Mathematics* Cambridge University Press, Cambridge.
- Atherton, R. W.; Schainker, R. B.; Ducot, E. R. (1975):** On the Statistical Sensitivity Analysis of Models for Chemical Kinetics. *AIChE Journal* **21**(3), 441.
- Babuška, R.; Verbruggen, H. B.; van Can H. J. L. (1999):** Fuzzy modeling of enzymatic penicillin-G conversion. *Engineering Applications of Artificial Intelligence* **12**(1), 79–92.

- Bailey, J. E. (1998):** Mathematical Modeling and Analysis in Biochemical Engineering: Past Accomplishments and Future Opportunities. *Biotechnology Progress* **14**(1), 8–20.
- Bakker, B. M.; Michels, P. A. M.; Oppendoes, F. R.; Westerhoff, H. V. (1997):** Glycolysis in Bloodstream Form *Trypanosoma brucei* Can Be Understood in Terms of the Kinetics of the Glycolytic Enzymes. *The Journal of Biological Chemistry* **272**(6), 3207–3215.
- Balci, O. (1997):** Verification, Validation and Accreditation of Simulation Models. In: Andradóttir, S.; Healy, K. J.; Withers, D.H.; Nelson, B.L. (Hrsg.): *Proceedings of the 1997 Winter Simulation Conference*, 135–141, New York, NY, USA, ACM Press.
- Balci, O. (1998):** Verification, Validation, and Testing. In: Banks, J. (Hrsg.): *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, Kap. 10, 335–393, John Wiley & Sons, New York.
- Balci, O.; Sargent, R. G. (1982a):** Some Examples of Simulation Model Validation Using Hypothesis Testing. In: Highland, H. J.; Chao, Y. W.; Madrigal, O. (Hrsg.): *Proceedings of the 1982 Winter Simulation Conference*, 620–629, New York, NY, USA, ACM Press.
- Balci, O.; Sargent, R. G. (1982b):** Validation of multivariate response simulation models by using hotelling's two-sample T^2 test. *Simulation* **39**(1-2), 185–192.
- Balci, O.; Sargent, R. G. (1984):** Validation of Simulation Models via Simultaneous Confidence Intervals. *American Journal of Mathematical and Management Science* **4**(3), 375–406.
- Baltes, M.; Schneider, R.; Sturm, C.; Reuss, M. (1994):** Optimal Experimental Design for Parameter Estimation in Unstructured Growth Models. *Biotechnology Progress* **10**(5), 480–488.
- Banks, J.; Carson II, J. S.; Nelson, B. L. (1995):** *Discrete-Event System Simulation*. Prentice Hall, Englewood Cliffs, N.J.
- Barabasi, A.; Albert, R. (1999):** Emergence of Scaling in Random Networks. *Science* **286**(5439), 509–512.
- Bäck, T.; Schwefel, H. P. (1993):** An Overview of Evolutionary Algorithms for Parameter Optimization. *Evolutionary Computation* **1**(1), 1–23.
- Bisswanger, H. (2000):** *Enzymkinetik: Theorie und Methoden*. Wiley VCH, Weinheim, 3 Aufl.
- Blangy, D.; Buc, H.; Monod, J. (1968):** Kinetics of the Allosteric Interactions of Phosphofructokinase from *Escherichia coli*. *Journal of Molecular Biology* **31**, 13–35.
- Blattner, F. R.; Plunkett III, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. (1997):** The Complete Genome Sequence of *Escherichia coli* K-12. *Science* **277**(5333), 1453–1462.
- Bock, H. G. (1981):** Numerical treatment of inverse problems in chemical reaction kinetics. In: Ebert, K.; Deuffhard, P.; Jäger, W. (Hrsg.): *Modelling of Chemical Reaction Systems*, 102–125, Springer-Verlag, Berlin.
- Bock, H. G. (1983):** Recent Advances in Parameter Identification Techniques for O.D.E. In: Deuffhard, P.; Hairer, E. (Hrsg.): *Numerical Treatment of Inverse Problems in Differential and Integral Equations*, Bd. 2 d. Reihe *Progress in Scientific Computing*, 95–121, Boston, Birkhäuser.
- Bodenstein, M. (1913):** Eine Theorie der photochemischen Reaktionsgeschwindigkeiten. *Zeitung für Physikalische Chemie* **85**, 329–368.

- Boiteux, A.; Goldbeter, A.; Hess, B. (1975):** Control of oscillating glycolysis of yeast by stochastic, periodic, and steady state source of substrate: model and experimental study. *Proceedings of the National Academy of Sciences of the USA* **72**(10), 3829–3833.
- Borg-Olivier, S. A.; Tarlinton, D.; Brown, K.D. (1987):** Defective Regulation of the Phenylalanine Biosynthetic Operon in Mutants of the Phenylalanyl tRNA Synthetase Operon. *Journal of Bacteriology* **169**(5), 1949–1953.
- Hanke Bourgeois, M. (2002):** *Grundlagen der Numerischen Mathematik und des Wissenschaftlichen Rechnens*. Teubner, Stuttgart, Leipzig, Wiesbaden.
- Box, M. J. (1966):** A comparison of several current optimization methods, and the use of transformations in constrained problems. *The Computer Journal* **9**(1), 67–77.
- Bronstein, I. N.; Semendajew, K. A. (1989):** *Taschenbuch der Mathematik*. Teubner Verlagsgesellschaft, Leipzig.
- Brown, K.D.; Sommerville, R. L. (1971):** Repression of Aromatic Amino Acid Biosynthesis in *Escherichia coli* K-12. *Journal of Bacteriology* **108**(1), 386–399.
- Brown, T. A. (2002):** *Genomes*. Wiley-Liss., New York, 2 Aufl.
- Broyden, C. G. (1970):** The convergence of a class of double-rank minimization algorithms, II: The new algorithm. *Journal of the Institute for Mathematics and Applications* **6**(1), 221–231.
- Buchholz, A. (1998):** Reaktionstechnische Untersuchungen zur Dynamik intrazellulärer Metabolite in glycerollimitierten *Escherichia Coli* K12. Diplomarbeit, Universität Bonn.
- Buchholz, A. (2001):** Quantifizierung intrazellulärer Metabolitdynamiken zur Untersuchung mikrobieller Stoffwechselnetzwerke. Dissertation, Universität Bonn.
- Buchholz, A.; Hurlebaus, J.; Wandrey, C.; Takors, R. (2002):** Metabolomics: quantification of intracellular metabolite dynamics. *Biomolecular Engineering* **19**(1), 5–15.
- Buchholz, A.; Takors, R.; Wandrey, C. (2001):** Quantification of Intracellular Metabolites in *Escherichia coli* K12 Using Liquid Chromatographic-Electrospray Ionization Tandem Mass Spectrometric Techniques. *Analytical Biochemistry* **295**(2), 129–137.
- Buckley, M. J. (1991):** Detecting a smooth signal: optimality of cusum based procedures. *Biometrika* **78**(2), 253–262.
- Burnham, K. P.; Anderson, D. R. (1998):** *Model Selection and Inference – A Practical Information-Theoretic Approach*. Springer-Verlag, New York, NY, USA, 2. Aufl.
- Burns, J. A.; Cornish-Bowden, A.; Groen, A. K.; Heinrich, R.; Kacser, H.; Porteous, J. W.; Rapoport, S. M.; Rapoport, T. A.; Stucki, J.W.; Tager, J. M.; Wanders, R. J. A.; Westerhoff, H. V. (1985):** Control of metabolic systems. *Trends in Biochemical Sciences* **10**(1), 16.
- Butcher, J. C. (1963):** Coefficients for the study of Runge-Kutta integration processes. *Journal of the Australian Mathematical Society* **3**, 185–201.
- Buziol, S.; Bashir, I.; Baumeister, A.; Claaßen, W.; Noisommit-Rizzi, N.; Mailinger, W.; Reuss, M. (2002):** New Bioreactor-Coupled Rapid Stopped-Flow Sampling Technique for Measurements of Metabolite Dynamics on a Subsecond Time Scale. *Biotechnology and Bioengineering* **80**(6), 632–636.
- Chain, S. J.; Chau, P. C. (1998):** Transition Probability Cell Cycle Model with Product Formation. *Biotechnology and Bioengineering* **58**(4), 387–399.

- Chance, B.; Garfinkel, D.; Higgins, J.; Hess, B. (1960):** Metabolic Control Mechanisms V. A SOLUTION FOR THE EQUATIONS REPRESENTING INTERACTION BETWEEN GLYCOLYSIS AND RESPIRATION IN ASCITES TUMOR CELLS. *The Journal of Biological Chemistry* **235**(8), 2426–2439.
- Chappell, M. J.; Godfrey, K. R. (1990):** Global identifiability of the parameters of nonlinear systems with specified inputs: A comparison of methods. *Mathematical Biosciences* **102**(1), 41–73.
- Chassagnole, C.; Noisommit-Rizzi, N.; Schmid, J.W.; Mauch, K.; Reuss, M. (2002):** Dynamic Modeling of the Central Carbon Metabolism of *Escherichia coli*. *Biotechnology and Bioengineering* **79**(1), 53–73.
- Chevalier, T.; Schreiber, I.; Ross, J. (1993):** Toward a Systematic Determination of Complex Reaction Mechanisms. *Journal of Physical Chemistry* **97**(26), 6776–6787.
- Christensen, B.; Nielsen, J. (1999):** Isotopomer Analysis Using GC-MS. *Metabolic Engineering* **1**(4), 282–290.
- Christensen, B.; Nielsen, J. (2000):** Metabolic Network Analysis: A Powerful Tool in Metabolic Engineering. *Advances in Biochemical Engineering/Biotechnology* **66**, 209–231.
- Clark-Curtiss, J. E.; Curtiss III, R. (1983):** Analysis of recombinant DNA using *Escherichia coli* minicells. *Methods in Enzymology* **101**, 347–362.
- Complex Biological Systems Initiatives (USA) (2004):** <http://www.nigms.nih.gov/funding/complex-systems.html>. Zugriffsdatum: 03/2004.
- Cornish-Bowden, A. (2004):** *Fundamentals of Enzyme Kinetics*. Portland Press, London, 3 Aufl.
- Cortassa, S.; Aon, M. A. (1994):** Metabolic control analysis of glycolysis and branching to ethanol production in chemostat cultures of *Saccharomyces cerevisiae* under carbon, nitrogen, or phosphate limitations. *Enzyme and Microbial Technology* **16**(9), 761–770.
- Cox, D.; Koh, E.; Wahba, G.; Yandell, B. S. (1988):** Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *The Annals of Statistics* **16**(1), 113–119.
- Curtiss, C. F.; Hirschfelder, J. O. (1952):** Integration of stiff equations. *Proceedings of the National Academy of Sciences of the USA* **38**, 235–243.
- Dahlquist, G. (1956):** Convergence and stability in the numerical integration of ordinary differential equations. *Mathematica Scandinavica* **4**, 33–53.
- van Dam, J.C.; Eman, M.R.; Frank, J.; Lange, H.C; van Dedem, G.W.K.; Heijnen, S.J. (2002):** Analysis of glycolytic intermediates in *Saccharomyces cerevisiae* using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. *Analytica Chimica Acta* **460**(2), 209–218.
- Dano, S.; Hynne, F.; De Monte, S.; d'Ovidio, F.; Sorensen, P. G.; Westerhoff, H. (2001):** Synchronization of glycolytic oscillations in a yeast cell population. *Faraday Discussions* **120**, 261–276.
- Davidson, F. A.; Xu, R.; Liu, J. (2002):** Existence and uniqueness of limit cycles in an enzyme-catalysed reaction system. *Applied mathematics and computation* **127**(2-3), 165–179.
- Degenring, D.; Froemel, C.; Dikta, G.; Takors, R. (2004a):** Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control* **14**(7), 729–745.

- Degenring, D.; Röhl, M.; Uhrmacher, A. M. (2003):** Discrete Event Simulation for a Better Understanding of Metabolite Channeling - A System Theoretic Approach. In: Priami, C. (Hrsg.): *Proceedings of the International Workshop on Computational Methods in Systems Biology*, Bd.2602 d. Reihe *Lecture Notes in Computer Science*, 114–126, Berlin- Heidelberg- New York, Springer-Verlag. ISBN 3-540-00605-2.
- Degenring, D.; Röhl, M.; Uhrmacher, A. M. (2004b):** Discrete Event, Multi-Level Simulation of Metabolite Channeling. *BioSystems* **75**(1-3), 29–41.
- Degenring, D.; Schröder, I.; Wandrey, C.; Liese, A.; Greiner, L. (2004c):** Resolution of 1,2-Diols by Enzyme-Catalyzed Oxidation with Anodic, Mediated Cofactor Regeneration in the Extractive Membrane Reactor: Gaining Insight by Adaptive Simulation. *Organic Research and Development* **8**(2), 213–218.
- Degenring, D.; Takors, R. (2002):** *In-vivo* Investigation of Metabolic Dynamics in *Escherichia Coli* K12 by Evaluation of Rapid Sampling Experiments. In: Polani, D.; Kim, J.; Martinetz, T. (Hrsg.): *Proceedings of 5th German Workshop on Artificial Life*, 27–33, Berlin, Akademische Verlagsgesellschaft Aka. ISBN 3-89838-030-0.
- Degenring, D.; Takors, R. (2003):** Modeling of Metabolism Dynamics in *Escherichia Coli* after Glucose Pulse Experiments. In: Troch, I.; Breitenecker, F. (Hrsg.): *Proceedings of 4th IMACS Symposium on Mathematical Modeling*, Bd.24(2), 368–374, Wien, ARGESIM-Verlag. ISBN 3-901608-24-9.
- delaFuente, I. M.; Martinez, L.; Veguillas, J.; Aguirregabiria, J. M. (1996):** Quasiperiodicity route to chaos in a biochemical system. *Biophysical Journal* **71**(5), 2375–2379.
- Delgado, J.; Liao, J. C. (1995):** Control of metabolic pathways by time-scale separation. *BioSystems* **36**(1), 55–70.
- Deuflhard, P.; Bornemann, F. (2002):** *Numerische Mathematik II – Gewöhnliche Differentialgleichungen*. de Gruyter, Berlin; New York, 2 Aufl.
- Deville-Bonne, D.; Garel, J. R. (1992):** A Conformational Transition Involved in Antagonistic Substrate Binding to the Allosteric Phosphofructokinase from *Escherichia coli*. *Biochemistry* **31**(6), 1695–1700.
- D’haeseleer, P.; Liang, S.; Somogyi, R. (2000):** Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* **16**(8), 707–726.
- Dickinson, R. P.; Gelin, R. J. (1976):** Sensitivity Analysis of Ordinary Differential Equation Systems: A Direct Method. *Journal of Computational Physics* **21**, 123–143.
- DIN 32 645 (1994):** Chemische Analytik, Nachweis-, Erfassungs- und Bestimmungsgrenze, Ermittlung unter Wiederholungsbedingungen, Begriffe, Verfahren, Auswertung.
- DIN 38 402 Teil 51 (1986):** Deutsche Einheitsverfahren zur Wasser, Abwasser und Schlammuntersuchung, Allgemeine Angaben (Gruppe A), Kalibrierung von Analysenverfahren, Auswertung von Meßergebnissen und lineare Kalibrierverfahren für die Bestimmung von Verfahrenskenngrößen (A 51).
- Domach, M. M.; Leung, S. K.; Cahn, R. E.; Cocks, G. G.; Shuler, M. L. (1984):** Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-A. *Biotechnology and Bioengineering* **26**(3), 203–216.
- van Domselaar, B.; Hemker, P. W. (1975):** Nonlinear parameter estimation in initial value problems. Technischer Bericht NW 18/75, Mathematisches Zentrum, Amsterdam.

- Dopheide, T. A. A.; Crewther, P.; Davidson, B. E. (1972):** Chorismate Mutase-Prephenate Dehydratase from *Escherichia coli* K-12, II. Kinetic Properties. *The Journal of Biological Chemistry* **247**(14), 4447–4452.
- Dougherty, E. P.; Hwang, J. T.; Rabitz, H. (1979):** Further developments and applications of the Green's function method of sensitivity analysis in chemical kinetics. *The Journal of Chemical Physics* **71**(4), 1794–1808.
- Dunker, A. M. (1981):** Efficient calculation of sensitivity coefficients for complex atmospheric models. *Atmospheric Environment* **15**(7), 1155–1161.
- Dunker, A. M. (1984):** The decoupled direct method for calculation of sensitivity coefficients in chemical-kinetics. *The Journal of Chemical Physics* **81**(5), 2385–2393.
- Dunn, I.; Heinzle, E.; Ingham, J.; Prenosil, J. (2003):** *Biological Reaction Engineering. Dynamic Modelling Fundamentals with Simulation Examples*. VCH Verlagsgesellschaft, Weinheim, 2 Aufl.
- Edwards, J. S.; Palsson, B. O. (2000a):** Robustness Analysis of the *Escherichia coli* Metabolic Network. *Biotechnology Progress* **16**(6), 927–939.
- Edwards, J. S.; Palsson, B. O. (2000b):** The *Escherichia coli* MG1655 *in silico* metabolic phenotype: Its definition, characteristics and capabilities. *Proceedings of the National Academy of Sciences of the USA* **97**(10), 5528–5533.
- Efron, B. (1979):** Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7**(1), 1–26.
- Eisen, M. B.; Spellman, P. T.; Brown, P. O.; Botstein, D. (1998):** Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the USA* **95**(25), 14863–14868.
- Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. (1997):** Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *Journal of Computer-Aided Molecular Design* **11**(5), 425–445.
- Ely, B.; Pittard, J. (1979):** Aromatic Amino Acid Biosynthesis: Regulation of Shikimate Kinase in *Escherichia coli* K-12. *Journal of Bacteriology* **138**(3), 933–943.
- Escherich, T. (1885):** Die Darmbakterien des Neugeborenen und Säuglings. *Fortschritte der Medizin* **3**, 515–522.
- Eubank, R. L.; Hart, J. D. (1993):** Commonality of cusum, von Neumann and smoothing-based goodness-of-fit tests. *Biometrika* **80**, 89–98.
- Eubank, R. L.; Spiegelman, C. H. (1990):** Testing the goodness-of-fit of a linear model via nonparametric regression techniques. *Journal of the American Statistical Society* **85**(410), 387–392.
- Falls, A. H.; McRae, G. J.; Seinfeld, J. H. (1979):** Sensitivity and uncertainty analysis of reaction mechanisms for photochemical air pollution. *International Journal of Chemical Kinetics* **10**, 1137–1162.
- Fehlberg, E. (1969):** Low-order classical Runge-Kutta formulas with stepsize control and their application to some heat transfer problems. Technischer Report R-135, NASA.
- Fell, D. (2003):** *Understanding the Control of Metabolism.*, Bd. 2 d. Reihe *Frontiers in Metabolism* Portland Press, London, Miami.
- Fell, D. A.; Wagner, A. (2000):** The small world inside large metabolic networks. *Nature Biotechnology* **18**(11), 1121–1122.

- Ferenci, T. (1996):** Adaption to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. *FEMS Microbiology Reviews* **18**(4), 301–317.
- Fiehn, O. (2001):** Combining genomics, metabolome analysis and biochemical modelling to understand metabolite networks. *Comparative and Functional Genomics* **2**(3), 155–168.
- Fiehn, O.; Kopka, J.; Dörrmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. (2000):** Metabolite profiling for plant functional genomics. *Nature Biotechnology* **18**(11), 1157–1161.
- Finney, A.; Hucka, M. (2003):** Systems biology markup language: Level 2 and beyond. *Biochemical Society Transactions* **31**(6), 1472–1473.
- Fischer, E. (1894):** . *Berichte der Deutschen Chemischen Gesellschaft* **27**, 2985–2993.
- Fisher, R. A. (1925):** Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society* **22**, 700–725.
- Fishman, G. S.; Kiviat, P. J. (1968):** The Statistics of Discrete-Event Simulation. *Simulation* **10**, 185–195.
- Fletcher, R. (1987):** *Practical Methods of Optimization*. John Wiley & Sons, New York, NY, USA, 2 Aufl.
- Fogel, L. J.; Owens, A. J.; Walsh, M. J. (1966):** *Artificial Intelligence Through Simulated Evolution*. John Wiley & Sons, New York, NY, USA.
- Friedman, N.; Linial, M.; Nachman, I.; Pe’er, D. (2000):** Using Bayesian networks to Analyze Expression Data. *Journal of Computational Biology* **3**/4, 601–620.
- Froemel, C. (2003):** Parameterreduktion in Stoffwechselmodellen mit Methoden der Statistik. Diplomarbeit, Fachhochschule Aachen, Abteilung Jülich.
- De la Fuente, A.; Brazhnik, P.; Mendes, P. (2001):** Two Complementary Methods for Reverse Engineering Gene Networks. In: Gauges, R.; van Gend, C.; Kummer, U. (Hrsg.): *Proceedings of the 2nd Workshop on Computation of Biochemical Pathways and Genetic Networks*, 45–53, Berlin, Logos Verlag.
- Gagneur, J.; Jackson, D. B.; Casari, G. (2003):** Hierarchical analysis of dependency in metabolic networks. *Bioinformatics* **19**(8), 1027–1034.
- Galazzo, J. L.; Bailey, J. E. (1990):** Fermentation pathway kinetics and metabolic flux control in suspended and immobilized *Saccharomyces cerevisiae*. *Enzyme and Microbial Technology* **12**(3), 162–172.
- Gancedo, J.M.; Gancedo, C. (1973):** Concentration of intermediary metabolites in yeast. *Biochemie* **55**, 205–211.
- Garcia-Olivares, A.; Villarroel, M.; Marijuan, P. C. (2000):** Enzymes as molecular automata: a stochastic model of self-oscillatory glycolytic cycles in cellular metabolism. *BioSystems* **56**(2-3), 121–129.
- Gautier, O.; Carr, R. W.; Seigneur, C. (1985):** Variational sensitivity analysis of a photochemical smog mechanism. *International Journal of Chemical Kinetics* **17**(12), 1347–1364.
- Genomes to Life (USA) (2004):** <http://doegenomestolife.org/>. Zugriffsdatum: 03/2004.
- Gerigk, M. (2001):** Integrierte Prozessentwicklung zur Herstellung von L-Phenylalanin mit *Escherichia coli*. Dissertation, Universität Berlin.

- Gibson, F (1964):** Chorismic acid: purification and some chemical and physical studies. *Biochemical Journal* **90**(2), 256–261.
- Gibson, F.; Pittard, J. (1968):** Pathways of biosynthesis of aromatic amino acids and vitamins and their control in microorganisms. *Bacteriological Reviews* **32**(4 Pt 2), 465–492.
- Gibson, M. A.; Bruck, J. (2000):** Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *Journal of Physical Chemistry A* **104**(9), 1876–1889.
- Giersch, C. (2003):** Stationary diffusion gradients associated with photosynthetic carbon flux – a study of compartmental versus diffusion-reaction models. *Journal of Theoretical Biology* **224**(3), 385–397.
- Gil, R.; Silva, F. J.; Zientz, E.; Delmotte, F.; González-Candelas, F.; Latorre, A.; Rausell, C.; Kamerbeek, J.; Gadau, J.; Hölldobler, B.; van Ham, R. C. H. J.; Gross, R.; Moya, A. (2003):** The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proceedings of the National Academy of Sciences of the USA* **100**(16), 9388–9393.
- Giuseppin, M. L.; van Riel, N. A. W. (2000):** Metabolic Modeling of *Saccharomyces cerevisiae* Using the Optimal Control of Homeostasis: A Cybernetic Model Definition. *Metabolic Engineering* **2**(1), 14–33.
- Giuseppin, M. L. F.; Verrips, C. T.; van Riel, N. A. W. (1999):** The cell factory needs a model of a factory. *Trends in Biotechnology* **17**(10), 383–384.
- Glaser, L.; Brown, D. H. (1955):** PURIFICATION AND PROPERTIES OF D-GLUCOSE-6-PHOSPHATE-DEHYDROGENASE. *The Journal of Biological Chemistry* **216**(1), 67–69.
- Glassbrock, N.; Beecher, C.; Ryals, C. (2000):** Metabolic profiling on the right path. *Nature Biotechnology* **18**(11), 1142–1143.
- Goldbeter, A.; Lefever, R. (1972):** Dissipative structures for an allosteric model: Application to glycolytic oscillations. *Biophysical Journal* **12**(10), 1302–1315.
- Gombert, A. K.; Nielsen, J. (2000):** Mathematical modelling of metabolism. *Current Opinion in Biotechnology* **11**(2), 180–186.
- Graf, U.; Henning, H. J.; Stange, K. (1966):** *Formeln und Tabellen der mathematischen Statistik*. Springer-Verlag, Berlin, Heidelberg, New York, 2. Aufl.
- Hairer, E.; Wanner, G. (1991):** *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. Springer Verlag, Berlin, Heidelberg, New York, 2 Aufl.
- Hartung, J.; Elpelt, B. (1995):** *Multivariate Statistik - Lehr- und Handbuch der angewandten Statistik*. R. Oldenbourg Verlag, München, 6 Aufl.
- Hartwell, L. H.; Hopfield, J. J.; Leibler, S.; Murray, A. W. (1999):** From molecular to modular cell biology. *Nature* **402**(6761), C47–C52.
- Hatzimanikatis, V. (1999):** Nonlinear metabolic control analysis. *Metabolic Engineering* **1**(1), 75–87.
- Hatzimanikatis, V.; Emmerling, M.; Sauer, U.; Bailey, J. E. (1998):** Application of mathematical tools for metabolic design of microbial ethanol production. *Biotechnology and Bioengineering* **58**(2-3), 154–161.
- Hatzimanikatis, V.; Lee, K. H. (1999):** Dynamical analysis of gene networks requires both mRNA and protein expression information. *Metabolic Engineering* **1**(4), 275–281.
- Haunschild, M. D. (2001):** Ein Werkzeug zur automatischen Modellgenerierung und verteilten Simulation metabolischer Netzwerke. Diplomarbeit, Universität Siegen.

- Haunschild, M. D.; Wiechert, W. (2003):** Sensitivity analysis of metabolic network models. In: Hohmann, R. (Hrsg.): *Proceedings of the 17th Simulation Symposium ASIM*, 415–420, Erlangen, San Diego, SCS Publishing House.
- Hauri, D. C.; Shen, P.; Arkin, A. P.; Ross, J. (1997):** Steady-State Measurements on the Fructose 6-Phosphate/Fructose 1,6-Bisphosphate Interconversion Cycle. *Journal of Physical Chemistry B* **101**(19), 3872–3876.
- Hedges Jr., R. M.; Rabitz, H. (1985):** Parametric sensitivity of system stability in chemical dynamics. *The Journal of Chemical Physics* **82**, 3674–3684.
- Heinrich, R.; Rapoport, T. A. (1974b):** A linear steady state treatment of enzymatic chains. Critique of the crossover theorem and a general procedure to identify interactions sites with an effector. *European Journal of Biochemistry* **42**(1), 97–105.
- Heinrich, R.; Rapoport, T. A. (1974a):** A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *European Journal of Biochemistry* **42**(1), 89–95.
- Heinrich, R.; Schuster, S. (1996):** *The Regulation Of Cellular Systems*. Kluwer Academic Publishers, Boston.
- Henri, V. (1902):** . *Comptes Rendus Hebdomadaires des Seances de l'Académie des Sciences* **135**, 916–919.
- Hermann, C. F. (1967):** Validation problems in games and simulations with special reference to models in international politics. *Behavioral Science* **12**, 216–231.
- Hewitt, C. J.; Nebe-Von-Caron, G.; Axelsson, B.; McFarlane, C. M.; Nienow, A. W. (2000):** Studies related to the scale-up of high-cell density *E. coli* fed-batch fermentations using multiparameter flow cytometry: Effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. *Biotechnology and Bioengineering* **70**(4), 381–390.
- Hill, A. V. (1910):** Possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *The Journal of Physiology* **40**, iv–vi.
- Hindmarsh, A. C. (1980):** LSODE and LSODI, two new initial value ordinary differential equation solvers. *ACM SIGNUM Newsletter* **15**, 10–11.
- Hirsch, E. (2001):** Reaktionstechnische Untersuchungen zur Auswirkung von Glucose-Pulsen auf den Aromabiosyntheseweg bei glucose-limitierten *E. coli* Stämmen. Diplomarbeit, Universität Flensburg.
- Hjemfelt, A.; Ross, J. (1995):** Implementation of logic functions and computations by chemical kinetics. *Physica D* **84**(1-2), 180–193.
- Hofmeyr, J. H. S. (1986):** Steady-state modeling of metabolic pathways: a guide for the prospective simulator. *Computer Applications in the Biosciences* **2**(1), 5–11.
- Holmberg, A. (1982):** On the Practical Identifiability of Microbial-Growth Models Incorporating Michaelis-Menten Type Nonlinearities. *Mathematical Biosciences* **62**(1), 23–43.
- Hood, L. (2001):** Unifying logic: Searching for the biggest truths in the smallest elements. http://www.geneforum.org/learnmore/articles/hoodl_200101.cfm.
- Hooke, R.; Jeeves, T. A. (1961):** 'Direct search' solution of numerical and statistical problems. *Journal of the Association for Computing Machinery* **8**(2), 212–229.

- Hoops, S.; de la Fuente, A.; Mendes, P. (2003):** Partial Estimation of the Jacobian from Time Course Data. In: Gauges, R.; Kummer, U.; Pahle, J.; Rost, U. (Hrsg.): *Proceedings of the 3rd Workshop on Computation of Biochemical Pathways and Genetic Networks*, 93–98, Berlin, Logos Verlag.
- Horbelt, W.; Müller, T.; Timmer, J.; Melzer, W.; Winkler, K. (2000):** Analysis of nonlinear differential equations: Parameter estimation and model selection. In: *Medical Data Analysis*, Bd.1933 d. Reihe *Lecture Notes in Computer Sciences*, 152–158, Berlin, Springer-Verlag.
- Horbelt, W.; Timmer, J.; Voss, H. U. (2002):** Parameter estimation in nonlinear delayed feedback systems from noisy data. *Physics Letters A* **299**(5-6), 513–521.
- Howden, W. E. (1980):** Functional program testing. *IEEE Transactions on Software Engineering* **6**(2), 162–169.
- Härdle, W.; Mammen, E. (1993):** Comparing Nonparametric Versus Parametric Regression Fits. *The Annals of Statistics* **21**(4), 1926–1947.
- Huang, X.; Holden, H. M.; Raushel, F. M. (2001):** Channeling of Substrates and Intermediates in Enzyme-Catalyzed Reactions. *Annual Review in Biochemistry* **70**, 149–180.
- Hucka, M.; Finney, A.; Sauro, H.; Bolouri, H. (2001):** Systems Biology Markup Language (SBML) Level 1: Structures and Facilities for Basic Model Definitions. Technischer Bericht MC 107-81, California Institute of Technology, Pasadena, CA 91125, USA.
- Hucka, M.; Finney, A.; Sauro, H. M.; Bolouri, H.; Doyle, J. C.; Kitano, H. (2003):** The Systems Biology Markup Language (SBML): A Medium for Representation and Exchange of Biochemical Network Models. *Bioinformatics* **19**(4), 524–531. <http://www.sbml.org>.
- Hurlebaus, J. (2001):** A Pathway Modeling Tool for Metabolic Engineering. Dissertation, Universität Bonn.
- Hurlebaus, J.; Buchholz, A.; Alt, W.; Wiechert, W.; Takors, R. (2002):** MMT - A pathway modeling tool for data from rapid sampling experiments. *In-Silico Biology* **2**(4), 467–484.
- Hwang, J. T. (1983):** Sensitivity analysis in chemical kinetics by the method of polynomial approximations. *International Journal of Chemical Kinetics* **15**(10), 959–987.
- Hwang, J. T.; Dougherty, E. P.; Rabitz, S.; Rabitz, H. (1978):** The Green's function method for sensitivity analysis in chemical kinetics. *The Journal of Chemical Physics* **69**(11), 5180–5191.
- Hynne, F.; Donø, S.; Sørensen, P. G. (2001):** Full-scale model of glycolysis in *Saccharomyces cerevisiae*. *Biophysical Chemistry* **94**(1-2), 121–163.
- Ideker, T.; Galitski, T.; Hood, L. (2001a):** A NEW APPROACH TO DECODING LIFE: Systems Biology. *Annual Review of Genomics and Human Genetics* **2**, 343–372.
- Ideker, T.; Thorsson, V.; Ranish, J. A.; Christmas, R.; Buhler, J.; Eng, J. K.; Bumgarner, R.; Goodlett, D. R.; Aebersold, R.; Hood, L. (2001b):** Integrated Genomic and Proteomic Analyses of a Systematically Perturbed Metabolic Network. *Science* **292**(5518), 929–934.
- Ingalls, B. P.; Sauro, H. M. (2002):** Sensitivity analysis of stoichiometric networks: an extension of metabolic control analysis to non-steady state trajectories. *Journal of Theoretical Biology* **222**(1), 23–36.
- Ingraham, J. L.; Maaløe, O.; Neidhardt, F. C. (1983):** *Growth of the Bacterial Cell*. Sinauer Associates Inc., Sunderland, Mass.

- Ishizuka, H.; Hanamura, A.; Kunimura, T.; Aiba, H. (1993):** A Lowered Concentration of cAMP-Receptor Protein Caused by Glucose is an Important Determinant for Catabolite Repression in *Escherichia coli*. *Molecular Microbiology* **10**(2), 341–350.
- Ito, T.; Chiba, T.; Ozawa, R.; Yoshida, M.; Hattori, M.; Sakaki, Y. (2001):** A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proceedings of the National Academy of Sciences of the USA* **98**(8), 4569–4574.
- Jeong, H.; Mason, S. P.; Barabasi, A. L.; Oltvai, Z. N. (2001):** Lethality and centrality in protein networks. *Nature* **411**(6833), 41–42.
- Jeong, H.; Tombor, B.; Albert, R.; Oltvai, Z. N.; Barabasi, A. L. (2000):** The large-scale organization of metabolic networks. *Nature* **407**(6804), 651–654.
- Johnson, J. L.; Reinhardt, G. D. (1994):** Influence of Substrates and MgADP on the Time-Resolved Intrinsic Fluorescence of Phosphofructokinase from *Escherichia coli*. Correlation of Tryptophan Dynamics to Coupling Entropy. *Biochemistry* **33**(9), 2644–2650.
- Johnson, J. L.; Reinhart, G. D. (1997):** Failure of a Two-State Model to Describe the Influence of Phospho(*enol*)pyruvate on Phosphofructokinase from *Escherichia coli*. *Biochemistry* **36**(42), 12814–12822.
- Johnson, R. A.; Wichern, D. W. (2002):** *Applied Multivariate Statistical Analysis*. Prentice Hall Inc., Englewood Cliffs, N.J., 5 Aufl.
- Jolliffe, I. T. (1972):** Discarding Variables in Principal Component Analysis. II: Real Data. *Applied Statistics* **22**, 21–31.
- Jolliffe, I. T. (1972):** Discarding Variables in Principal Component Analysis. I: Artificial Data. *Applied Statistics* **21**, 160–173.
- Jones, M. E.; Berry, M. N.; Phillips, J. W. (2002):** Futile Cycles Revisited: A Markov Chain Model of Simultaneous Glycolysis and Gluconeogenesis. *Journal of Theoretical Biology* **217**(4), 509–523.
- de Jong, H. (2000):** Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. Technischer Bericht 4032, INRIA Rhône-Alpes.
- Jonker, C. M.; Snoep, J. L.; Treur, J.; Westerhoff, H. V.; Wijngards, W. C. A (2002):** Putting Intentions into Cell Biochemistry: An Artificial Intelligence Perspective. *Journal of Theoretical Biology* **214**(1), 105–134.
- Julien, S.; Babary, J. P.; Lessard, P. (1998):** Theoretical and practical identifiability of a reduced order model in an activated sludge process doing nitrification and denitrification. *Water Science and Technology* **37**(12), 309–316.
- Kacser, H.; Burns, J. A. (1973):** The control of flux. *Symposium of the Society for Experimental Biology* **27**, 65–104.
- Kanehisa, M.; Goto, S. (2000):** KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**(1), 27–30. <http://www.genome.ad.jp/kegg>.
- Karp, P. D.; Riley, M.; Saier, M.; Paulsen, I. T.; Paley, S. M.; Pellegini-Toole, A. (2000):** The EcoCyc and MetaCyc databases. *Nucleic Acids Research* **28**, 56–59. <http://ecocyc.org>.
- Kauffman, K. J.; Pajerowski, J. D.; Jamshidi, N.; Palsson, B. O.; Edwards, J. S. (2002):** Description and Analysis of Metabolic Connectivity and Dynamics in the Human Red Blood Cell. *Biophysical Journal* **83**(2), 646–662.

- Kennerknecht, N.; Peters-Wendisch, P.; Eggeling, L.; Sahm, H. (2003):** Metabolic Engineering: Entwicklung von Bakterienstämmen für die Lysinproduktion. *BIOspektrum* **9**, 582–585.
- Küffner, R.; Zimmer, R.; Lengauer, T. (2000):** Pathway analysis in metabolic databasis via differential metabolic display (DMD). *Bioinformatics* **16**(9), 825–836.
- Khanna, S. (1991):** Logic Programming for Software Verification and Testing. *The Computer Journal* **34**(4), 350–357.
- Kholodenko, B. N.; Demin, O. V.; Westerhoff, H. V. (1997):** Control Analysis of Periodic Phenomena in Biological Systems. *Journal of Physical Chemistry B* **101**(11), 2070–2081.
- Kierzik, A. M. (2002):** STOCKS: STOChastic Kinetic Simulation of biochemical systems with Gillespie algorithm. *Bioinformatics* **18**(3), 470–481.
- King, E. L.; Altmann, C. (1956):** A schematic method of deriving rate laws for enzyme-catalyzed reactions. *The Journal of Physical Chemistry B* **60**, 1375–1378.
- Kinnebrock, W. (1994):** *Optimierung mit genetischen und selektiven Algorithmen*. Oldenbourg Verlag, München, Wien.
- Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. (1983):** Optimization by simulated annealing. *Science* **220**(4598), 671–680.
- Kitano, H. (2000):** Perspectives on Systems Biology. *New Generation Computing* **18**(3), 199–216.
- Kitano, H. (2002a):** Computational systems biology. *Nature* **420**(6912), 206–210.
- Kitano, H. (2002b):** Systems Biology: A Brief Overview. *Science* **295**(5560), 1662–1664.
- Kohn, M. C.; Lemieux, D. R. (1991):** Identification of regulatory properties of metabolic networks by graph theoretical modeling. *Journal of Theoretical Biology* **150**(1), 3–25.
- Kolmogorov, A. N. (1941):** Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics* **12**, 461–463.
- de Koning, W.; van Dam, K. (1992):** A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Analytical Biochemistry* **204**(1), 118–123.
- Kopperschlager, G.; Augustin, H.W. (1967):** Fehlermöglichkeiten bei der Bestimmung von Metabolitgehalten in Hefezellen. *Experientia* **23**(8), 623–624.
- Koshland Jr., D. E.; Nementhy, G.; Filmer, D. (1966):** Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **5**(1), 365–385.
- Kosmol, P. (1993):** *Methoden zur numerischen Behandlung nichtlinearer Gleichungen und Optimierungsaufgaben*. Teubner, Stuttgart.
- Kramer, M. A.; Calo, J. M.; Rabitz, H. (1981):** An improved computational method for sensitivity analysis - Green's-function method with aim. *Applied Mathematical Modelling* **5**(6), 432–441.
- Kreft, J. U.; Booth, G.; Winpenney, J. W. T. (1998):** BacSim, a simulator for individual-based modelling of bacterial colony growth. *Microbiology* **144**(12), 3275–3287.
- Kremling, A. (2001):** Strukturierung zellulärer Funktionseinheiten - ein signalorientierter Modellierungsansatz für zelluläre Systeme am Beispiel von *Escherichia coli*. Dissertation, Universität Stuttgart.

- Kremling, A.; Bettenbrock, K.; Laube, B.; Jahreis, K.; Lengeler, J. W.; Gilles, E. D. (2001):** The Organization of Metabolic Reaction Networks: III. Application for Diauxic Growth on Glucose and Lactose. *Metabolic Engineering* **3**(4), 362–379.
- Kremling, A.; Gilles, E. D. (2001):** The Organization of Metabolic Reaction Networks II. Signal Processing in Hierarchical Structured Functional Units. *Metabolic Engineering* **3**(2), 138–150.
- Kremling, A.; Jahreis, K.; Lengeler, J. W.; Gilles, E. D. (2000):** The Organization of Metabolic Reaction Networks: A Signal Oriented Approach to Cellular Models. *Metabolic Engineering* **2**(3), 190–200.
- Krogh, F. T. (1974):** Changing step size in the integration of differential equations using modified divided differences. In: *Proceedings of the Conference on Numerical ordinary differential equations*, Bd.362 d. Reihe *Lecture Notes in Mathematics*, 22–71, Berlin, Heidelberg, New York, Springer-Verlag.
- Kruger, N. J. (1995):** ERRORS AND ARTEFACTS IN COUPLED SPECTROPHOTOMETRIC ASSAYS OF ENZYME ACTIVITY. *Phytochemistry* **38**(5), 1065–1071.
- Kundig, W.; Gosh, S.; Roseman, S. (1964):** Phosphate bound to histidine in a protein as an intermediate in a novel phospho-transferase system. *Proceedings of the National Academy of Sciences, USA* **52**, 1067–1074.
- Kuo, D.; Keasling, J. D. (1996):** A Monte Carlo simulation of plasmid replication during the bacterial division cycle. *Biotechnology and Bioengineering* **52**(6), 633–647.
- Kutta, W. (1901):** Beitrag zur näherungsweise Integration totaler Differentialgleichungen. *Zeitschrift für Mathematik und Physik* **46**, 436–453.
- Laffend, L.; Shuler, M. L. (1994):** Ribosomal-protein limitations in *Escherichia-coli* under conditions of high translational activity. *Biotechnology and Bioengineering* **43**(5), 388–398.
- Lagarius, J. C.; Reeds, J. A.; Wright, M. H.; Wright, P. E. (1998):** Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization* **9**(1), 112–147.
- Lange, H. C.; Eman, M.; van Zuijlen, G.; Visser, D.; van Dam, J. C.; Frank, J.; Teixeira de Mattos, M. J.; Heijnen, J. J. (2001):** Improved Rapid Sampling for in vivo Kinetics of Intracellular Metabolites in *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering* **75**(4), 406–415.
- Larter, R.; Rabitz, H.; Kramer, M. (1984):** Sensitivity analysis of limit cycles with application to the Brusselator. *The Journal of Chemical Physics* **80**(9), 4120–4128.
- Law, A. M.; Kelton, W. D. (1991):** Building valid and credible simulation models. In: Munson, E. M.; Luhrs, M. (Hrsg.): *Simulation Modeling & Analysis*, Kap. 5, 298–324, McGraw-Hill Inc., New York.
- Lee, B.; Yen, J.; Yang, L.; Liao, J. C. (1999):** Incorporating Qualitative Knowledge in Enzyme Kinetic Models Using Fuzzy Logic. *Biotechnology and Bioengineering* **62**(6), 722–729.
- Lee, K. H. (2001):** Proteomics: a technology-driven and technology-limited discovery science. *Trends in Biotechnology* **19**(6), 217–222.
- Lehninger, A. L.; Nelson, D. L.; Cox, M. M. (1994):** *Prinzipien der Biochemie*. Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford, 2 Aufl.
- Lemmen, C.; Lengauer, T. (1997):** Time-efficient flexible superposition of medium-sized molecules. *Journal of Computer-Aided Molecular Design* **11**(4), 357–368.
- Lengeler, J.; Drews, G.; Schlegel, H. G. (1999):** *Biology of the Prokaryotes*. Blackwell Science Inc., Malden, USA.

- Lengeler, J. W. (1993):** Carbohydrate transport in bacteria under environmental conditions, a black box? *Antonie van Leeuwenhoek* **63**(3-4), 275–288.
- Levenberg, K. (1944):** A method for the solution of certain nonlinear problems in least squares. *Quarterly Journal on Applied Mathematics* **2**(2), 164–168.
- Lewin, B. (2000):** *Genes VII*. Oxford University Press, Inc., New York, USA, 7 Aufl.
- Li, Z.; Laidig, K. E.; Dagett, V. (1998):** Conformational Search Using a Molecular Dynamics-Minimization Procedure: Applications to Clusters of Coulombic Charges, Lennard-Jones Particles, and Waters. *Journal of Computational Chemistry* **19**(1), 60–70.
- Liang, S.; Fuhrman, S.; Somogyi, R. (1998):** Reveal, a general reverse engineering algorithm for inference of genetic network architectures. *Pacific Symposium of Biocomputing* **3**, 18–29.
- Lin, H. Y.; Mathisizik, B.; Xu, B.; Enfors, S. O.; Neubauer, P. (2001):** Determination of the Maximum Specific Uptake Capacities for Glucose and Oxygen in Glucose-Limited Fed-Batch Cultivations of *Escherichia Coli*. *Biotechnology and Bioengineering* **73**(5), 347–357.
- Lineweaver, H.; Burk, D. (1934):** The determination of enzyme dissociation constants. *The Journal of the American Chemical Society* **56**, 658–666.
- Liu, R. Y. (1988):** Bootstrap Procedures Under Some non-i.i.d. Models. *The Annals of Statistics* **16**(4), 1696–1708.
- Ljung, L. (1999):** *System Identification: Theory for the User*. Prentice-Hall, Upper Saddle River, N.J., 2 Aufl.
- Ljung, L.; Glad, T. (1994a):** *Modeling of Dynamic Systems*. Prentice-Hall, Upper Saddle River, N.J.
- Ljung, L.; Glad, T. (1994b):** On global identifiability for arbitrary model parametrizations. *Automatica* **30**(2), 265–276.
- Lux, R.; Jahreis, K.; Bettenbrock, K.; Parkinson, J.S.; Lengeler, J.W. (1995):** Coupling the phosphotransferase system and the methyl-accepting chemotaxis protein-dependent chemotaxis signaling pathways of *Escherichia coli*. *Proceedings of the National Academy of Sciences, USA* **92**(25), 11583–11587.
- Madigan, M. T.; Martinko, J.; Parker, J. (2002):** *Brock Biology of Microorganisms*. Prentice Hall, Inc., Englewood Cliffs, N.J., 10 Aufl.
- Mailinger, W.; Baumeister, A.; Reuss, M.; Rizzi, M. (1998):** Rapid and highly automated determination of adenine and pyridine nucleotides in extracts of *Saccharomyces cerevisiae* using a micro robotic sample preparation-HPLC system. *Journal of Biotechnology* **63**(2), 155–157.
- Maitra, U. S.; Sprinson, D. B. (1978):** 5-Dehydro-3-Deoxy-D-arabino-heptulosonic-Acid-7-Phosphate. An Intermediate of the 3-Dehydroquinate Synthase Reaction. *The Journal of Biological Chemistry* **253**(15), 5426–5430.
- Makino, S.; Kuntz, I. D. (1997):** Automated Flexible Ligand Docking Method and Its Application for Database Search. *Journal of Computational Chemistry* **18**(14), 1812–1825.
- Malcovati, M.; Valentini, G. (1982):** AMP- and fructose 1,6-bisphosphate-activated pyruvate kinases from *Escherichia coli*. *Methods in Enzymology* **90**, 170–179.
- Mammen, E. (1992):** When Does Bootstrap Work? Asymptotic Results and Simulations. Nr. 77 in *Lecture Notes in Statistics*, Heidelberg, Berlin, New York, Springer-Verlag.

- Mandell, A.; Selz, K. A.; Shlesinger, M. F. (2000):** Predicting Peptide-Receptor, Peptide-Protein and Chaperone-Protein Binding Using Patterns in Amino Acid Hydrophobic Free Energy Sequences. *The Journal of Physical Chemistry B* **104**(16), 3953–3959.
- Mantzaris, N. V.; Liou, J. J.; Daoutidis, P.; Srieenc, F. (1999):** Numerical solution of a mass structured cell population balance model in an environment of changing substrate concentration. *Journal of Biotechnology* **71**(1-3), 157–174.
- Mardia, K. V.; Kent, J. T.; Bibby, J. M. (1979):** *Multivariate Analysis*. Academic Press, London.
- Marquardt, D. W. (1963):** An algorithm for least squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics* **11**(2), 431–441.
- Mauch, K.; Arnold, S.; Reuss, M. (1997):** Dynamic sensitivity analysis for metabolic systems. *Chemical Engineering Science* **52**(15), 2589–2598.
- McAdams, H. H.; Arkin, A. (1998):** SIMULATION OF PROKARYOTIC GENETIC CIRCUITS. *Annual Review of Biophysics and Biomolecular Structure* **27**, 199–224.
- McKinnon, K. I. M. (1998):** Convergence of the Nelder-Mead Simplex Method to a Nonstationary Point. *SIAM Journal on Optimization*, **9**(1), 148–158.
- Mendes, P.; Kell, D. B. (1998):** Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* **14**(10), 869–883.
- Michaelis, L.; Menten, M. L. (1913):** Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift* **49**, 333–369.
- Microbial Cell Project (USA) (2004):** <http://microbialcellproject.org/>. Zugriffsdatum: 03/2004.
- Miller, D. R. (1974):** Sensitivity analysis and validation of simulation models. *Journal of Theoretical Biology* **48**(2), 345–360.
- Minsky, M. (1968):** Matter, Mind and Models. In: Minsky, M. (Hrsg.): *Semantic information processing*, 425–432, Cambridge, Mass., MIT Press.
- Misset, O.; Blaauw, M.; Postma, P. W.; Robillard, G. T. (1983):** Bacterial phosphoenolpyruvat-dependent phosphotransferase system. Mechanism of the transmembrane sugar translocation and phosphorylation. *Biochemistry* **22**(26), 6163–6170.
- Müller, T. G.; Noykova, N.; Gyllenberg, M.; Timmer, J. (2002):** Parameter identification in dynamical models of anaerobic waste water treatment. *Mathematical Biosciences* **177 & 178**, 147–160.
- Molnar, M.; Vas, M. (1993):** Mg^{2+} affects the binding of ADP but not ATP to 3-phosphoglycerate kinase. Correlation between equilibrium dialysis binding and enzyme kinetic data. *Biochemical Journal* **293**(2), 595–599.
- Monod, J. (1942):** Recherches sur la Croissance des Cultures Bactériennes. Dissertation, Herman, Paris.
- Moritz, B.; Striegel, K.; de Graaf, A.A.; Sahm, H. (2000):** Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from *Corynebacterium glutamicum* and their application for predicting pentose phosphate pathway flux *in vivo*. *European Journal of Biochemistry* **267**(12), 3442–3452.
- Morton-Firth, C. J.; Bray, D. (1998):** Predicting Temporal Fluctuations in an Intracellular Signalling Pathway. *Journal of Theoretical Biology* **192**(1), 117–128.

- Munack, A. (1989):** Optimal feeding strategy for identification of Monod-type models by fed-batch experiments. In: et al., Fish N. M. (Hrsg.): *Computer Applications in Fermentation Technology: Modelling and Control of Biotechnological Processes*, 195–204, London, New York, Elsevier Applied Sciences.
- Munack, A. (1995):** Simulation bioverfahrenstechnischer Prozesse. In: Schuler, H. (Hrsg.): *Prozeßsimulation*, VCH, Weinheim, New York, Basel, Cambridge.
- Munack, A.; Posten, C. (1989):** Design of Optimal Dynamical Experiments for Parameter Estimation in Bioprocesses. In: *Proceedings of the American Control Conference ACC89*, 2011–2016, Pittsburgh, PA.
- Neijssel, O. M.; Hueting, S.; Tempest, D. W. (1977):** Glucose transport capacity is not the rate-limiting step in the growth of some wild-type strains of *Escherichia coli* and *Klebsiella aerogenes* in chemostat culture. *FEMS Microbiology Letters* **2**(1-3), 1–3.
- Nelder, J. A.; Mead, R. (1965):** A simplex method for function minimization. *The Computer Journal* **7**(4), 308–313.
- Nelles, O. (2001):** *Nonlinear System Identification*. Springer-Verlag, Berlin, Heidelberg, New York.
- Nielsen, J. (1997):** Metabolic control analysis of biochemical pathways based on a thermokinetic description of reaction rates. *Biochemical Journal* **321**(1), 133–138.
- Nielsen, J.; Jørgensen, H. S. (1995):** Metabolic Control Analysis of the Penicillin Biosynthetic Pathway in a High-Yielding Strain of *Penicillium chrysogenum*. *Biotechnology Progress* **11**(3), 299–305.
- Nielsen, J.; Jørgensen, H. S. (1996):** A KINETIC MODEL FOR THE PENICILLIN BIOSYNTHETIC PATHWAY IN *Penicillium chrysogenum*. *Control Engineering Practice* **4**(6), 765–771.
- Nielsen, K.; Sørensen, P. G.; Hynne, F.; Busse, H. G. (1998):** Sustained oscillations in glycolysis: an experimental and theoretical study of chaotic and complex periodic behavior and of quenching of simple oscillations. *Biophysical Chemistry* **72**(1-2), 49–62.
- Nihtilä, M.; Virkkunen, J. (1977):** Practical Identifiability of Growth and Substrate Consumption Models. *Biotechnology and Bioengineering* **19**(12), 1831–1850.
- Nissen, T. L.; Schulze, U.; Nielsen, J.; Villadsen, J. (1997):** Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*. *Microbiology* **143**(1), 203–218.
- Norden, R. H. (1972):** A survey of maximum likelihood estimation. I. *International Statistical Review* **40**(3), 329–354.
- Nordsieck, A. (1962):** On numerical integration of ordinary differential equations. *Mathematics of Computation* **16**, 22–49.
- Oberhammer, H. (1998):** Molecular Structures and Conformations: Experiment and Theory. *Journal of Computational Chemistry* **19**(2), 123–128.
- Oberle, H. J.; Sothman, B. (1999):** Numerical Computation of Optimal Feed Rates for a Fed-Batch Fermentation Model. *Journal of Optimization Theory and Applications* **100**(1), 1–13.
- Ogino, T.; Garner, C.; Markley, J. L.; Hermann, K. M. (1982):** Biosynthesis of Aromatic Compounds: ¹³C NMR Spectroscopy of whole *Escherichia coli* Cells. *Proceedings of the National Academy of Sciences, USA* **79**(19), 5828–5832.

- Oldiges, M. (2004):** Metabolomanalyse zur Untersuchung der Dynamik im Aromatenbiosyntheseweg in L-Phenylalanin Produzenten von *Escherichia coli*. Dissertation, Universität Bonn. (in Vorbereitung).
- Oldiges, M.; Takors, R. (2004):** Applying Metabolic Profiling Techniques for Stimulus-Response Experiments: Chances and Pitfalls. *Advances in Biochemical Engineering/Biotechnology*. (angenommen).
- Ostergaard, S.; Olsson, L.; Nielsen, J. (2001):** In Vivo Dynamics of Galactose Metabolism in *Saccharomyces Cerevisiae*: Metabolic Fluxes and Metabolite Levels. *Biotechnology and Bioengineering* **73**(5), 412–425.
- Page, B. (1991):** Modellvalidierung und -dokumentation. In: *Diskrete Simulation. Eine Einführung mit Modula-2*, Kap. 5, 145–156, Springer Verlag, Berlin.
- Paker, W. J. (1997):** A multi-parametric sensitivity analysis of a model describing the fate of volatile organic compounds in tricking filters. *Journal of the Air and Waste Management Association* **47**, 871–880.
- Palsson, B. O. (2000):** The challenges of in silico biology. *Nature Biotechnology* **18**(11), 1147–1150.
- Pandis, S. N.; Seinfeld, J. H. (1989):** Sensitivity analysis of a chemical mechanism for aqueous-phase atmospheric chemistry. *Journal of Geophysical Research* **94**(D1), 1105–1126.
- Park, J.; Lappe, M.; Teichman, S. A. (2001):** Mapping Protein Family Interactions: Intramolecular and Intermolecular Protein Family Interaction Repertoires in the PDB and Yeast. *Journal of Molecular Biology* **307**(3), 929–938.
- Paschold, H.; Weuster-Botz, D.; Schaefer, U.; Boos, W. (1998):** Verfahren zur schnellen Probennahme biologischer Proben. Technischer Bericht, *Patentschrift DE 197 05 289 C1*.
- Petersen, S. (2000):** Untersuchungen zur *in vivo* Aktivität anaplerotischer Stoffwechselwege in *Corynebacterium glutamicum* mittels ¹³C-Markierungstechnik. Dissertation, Heinrich-Heine-Universität Düsseldorf.
- Petzold, L. R. (1983):** Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. *Siam Journal of Scientific Computation* **4**(1), 136–148.
- Petzold, L. R.; Hindmarsh, A. C. (2004):** Livermore solver for ordinary differential equations. <http://www.netlib.org>, Zugriffsdatum: 03/2004.
- Pissara, P. N.; Nielsen, J.; Bazin, M. J. (1996):** Pathway kinetics and metabolic control analysis of a high-yielding strain of *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnology and Bioengineering* **51**(2), 168–176.
- Pittard, A.J. (1999):** Biosynthesis of the Aromatic Amino Acids. In: Neidhardt, F.C. (Hrsg.): *Escherichia Coli and Salmonella: Cellular and Molecular Biology*, Kap. 28, 458–485, ASM Press, Washington.
- Postma, P. W.; Lengeler, J. W.; Jacobson, G. R. (1993):** Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiology Reviews* **57**(3), 543–594.
- Postma, P.W.; Lengeler, J.W.; Jacobson, G.R. (1999):** Phosphoenolpyruvate:Carbohydrate Phosphotransferase Systems. In: Neidhardt, F.C. (Hrsg.): *Escherichia Coli and Salmonella: Cellular and Molecular Biology*, Kap. 75, 1149–1174, ASM Press, Washington.

- Pramanik, J.; Keasling, J. D. (1997):** Stoichiometric Model of *Escherichia coli* Metabolism: Incorporation of Growth-Rate Dependent Biomass Composition and Mechanistic Energy Requirements. *Biotechnology and Bioengineering* **56**(4), 398–421.
- Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterling, W. T. (1992):** *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 2 Aufl.
- Price, C. J.; Coope, I. D.; Byatt, D. (2002a):** A Convergent Variant of the Nelder-Mead Algorithm. *Journal of Optimization Theory and Applications* **113**(1), 5–19.
- Price, N. D.; Papin, J.A.; Palsson, B. O. (2002b):** Determination of Redundancy and Systems Properties of the Metabolic Network of *Helicobacter pylori* Using Genome-Scale Extreme Pathway Analysis. *Genome Research* **12**(5), 760–769.
- Prince, P. J.; Dormand, J. R. (1981):** High order embedded Runge-Kutta formulae. *Journal of Computational and Applied Mathematics* **7**(1), 67–75.
- Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, A.; Walsh, M. C.; Berden, J. A.; Brindle, K. M.; Kell, D. B.; Rowland, J. J.; Westerhoff, H. V.; van Dam, K.; Oliver, S. G. (2001):** A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutants. *Nature Biotechnology* **19**(1), 45–50.
- Rabitz, H.; Kramer, M.; Dacol, D. (1983):** Sensitivity Analysis in Chemical Kinetics. *Annual Review of Physical Chemistry* **34**, 419.
- Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. (1996):** A Fast Flexible Docking Method Using an Incremental Construction Algorithm. *Journal of Molecular Biology* **261**(3), 470–489.
- Reddy, V. N.; Liebman, M. N.; Mavrovouniotis, M. L. (1996):** Qualitative analysis of biochemical reaction systems. *Computers in Biology and Medicine* **26**(1), 9–24.
- Reder, C. (1988):** Metabolic control theory: a structural approach. *Journal of Theoretical Biology* **135**(2), 175–201.
- Richer, E.; Courville, P.; Bergevin, I.; Cellier, M. F. M. (2003):** Horizontal Gene Transfer of “Prototype” Nramp in Bacteria. *Journal of Molecular Evolution* **57**(4), 363–376.
- Richter, O.; Betz, A.; Giersch, C. (1975):** The response of oscillating glycolysis to perturbations in the NADH/NAD system: a comparison between experiments and a computer model. *BioSystems* **7**(1), 137–146.
- Rinnooy Kan, A. H. G.; Timmer, G. T. (1989):** Global optimization: a survey. In: Penot, J. P. (Hrsg.): *New methods in optimization and their industrial uses : state of the art, recent advances, perspectives*, Bd. 87 d. Reihe *International Series of Numerical Mathematics*, 133–155, Basel, Birkhäuser Verlag.
- Rizzi, M.; Theobald, U.; Querfurth, E.; Baltes, M.; Reuss, M. (1996):** In Vivo Investigations of Glucose Transport in *Saccharomyces Cerevisiae*. *Biotechnology and Bioengineering* **49**(3), 316–327.
- Roels, J. A. (1983):** *Energetics and Kinetics in Biotechnology*. Elsevier Biomedical Press, Amsterdam.
- Rohwer, J. M.; Meadow, N. D.; Roseman, S.; Westerhoff, H. V.; Postma, P. W. (2000):** Understanding Glucose Transport by the Bacterial Phosphoenolpyruvate:Glycose Phosphotransferase System on the Basis of Kinetic Measurements *in Vitro*. *The Journal of Biological Chemistry* **275**(10), 34909–34921.
- Rohwer, J. M.; Postma, P. W.; Kholodenko, B. N.; Westerhoff, H. V. (1998):** Implications of macromolecular crowding for signal transduction and metabolite channeling. *Proceedings of the National Academy of Sciences of the USA* **95**(18), 10457–10552.

- Rosenbrock, H. H. (1960):** An Automatic Method for finding the Greatest or Least Value of a Function. *The Computer Journal* **3**(3), 175–184.
- Ross, J. (2003):** New Approaches to the Deduction of Complex Reaction Mechanisms. *Accounts of Chemical Research* **36**(11), 839–847.
- Ross, J.; Vlad, M. O. (1999):** Nonlinear Kinetics and New Approaches to Complex Reaction Mechanisms. *Annual Review of Physical Chemistry* **50**, 51–78.
- Roux-Rouquié, M.; Hood, L.; Imbeaud, S.; Auffray, C. (2003):** Issues in Computational Methods for Functional Genomics and Systems Biology. In: Priami, C. (Hrsg.): *Computational Methods in Systems Biology*, Bd.2602 d. Reihe *Lecture Notes in Computer Science*, 182–186, Berlin, Heidelberg, New York, Springer-Verlag. ISBN 3-540-00605-2.
- Rowan, T. (1990):** Functional stability analysis of numerical algorithms. Dissertation, University of Texas at Austin.
- Runge, C. (1895):** Über die numerische Auflösung von Differentialgleichungen. *Mathematische Annalen* **46**, 167–178.
- Russell, S.; Norvig, P. (1995):** *Artificial Intelligence. A Modern Approach*. Prentice-Hall, Englewood Cliffs, N.J.
- Rzhetsky, A.; Gomez, S. M. (2001):** Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome. *Bioinformatics* **17**(10), 988–996.
- Sachs, L. (2004):** *Angewandte Statistik: Anwendung statistischer Methoden*. Springer-Verlag, Berlin, Heidelberg, New York, 11 Aufl.
- Saltelli, A.; Chan, K.; Scott, E. M.(Hrsg.) (2001):** *Sensitivity Analysis*. John Wiley & Sons, New York, Weinheim, Queensland.
- Samoilov, M.; Arkin, A.; Ross, J. (2001):** On the deduction of chemical reaction pathways from measurements of time series of concentrations. *Chaos* **11**(1), 108–114.
- Santillán, M.; Mackey, M. C. (2001):** Dynamic regulation of the tryptophan operon: A modeling study a comparison with experimental data. *Proceedings of the National Academy of Sciences of the USA* **98**(4), 1364–1369.
- Sargent, R. G. (1982):** Verification and Validation of Simulation Models. In: Cellier, F. E. (Hrsg.): *Progress in Modelling and Simulation*, Kap. 9, 159–169, Academic Press, London.
- Sargent, R. G. (1996):** Some Subjective Validation Methods Using Graphical Displays of Data. In: Charnes, J.; Morrice, D. J.; Brunner, D. T.; Swain, J. J. (Hrsg.): *Proceedings of the 1996 Winter Simulation Conference*, 345–351, New York, NY, USA, ACM Press.
- Sargent, R. G. (2000):** Verification, validation, and accreditation of simulation models. In: Fishwick, P. A.; Kang, K.; Joines, J. A.; Barton, R. R. (Hrsg.): *Proceedings of the 2000 Winter Simulation Conference*, 50–59, San Diego, CA, USA, Society for Computer Simulation International.
- Sargent, R. G. (2001):** Some approaches and paradigms for verifying and validating simulation models. In: Peters, B. A.; Smith, J. S.; Medeiros, D.; Rohrer, M. W. (Hrsg.): *Proceedings of the 2001 Winter Simulation Conference*, 106–114, Washington, DC, USA, IEEE Computer Society.
- Sauter, T.; Kremling, A.; Bettenbrock, K.; Fischer, S.; Gilles, E. D. (2002):** Analysis of the dynamics of the *Escherichia coli* glucose PTS in different time windows. In: Polani, D.; Kim, J.; Martinetz, T. (Hrsg.): *Fifth German Workshop on Artificial Life: Abstracting and Synthesizing the Principles of life*, 33–42, Leipzig, Akademische Verlagsgesellschaft Aka. ISBN 3-89838-030-0.

- Savageau, M. A. (1969a):** Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *Journal of Theoretical Biology* **25**(3), 365–369.
- Savageau, M. A. (1969b):** Biochemical systems analysis. II. The steady-state solutions for an n-pool system using a power-law approximation. *Journal of Theoretical Biology* **25**(3), 370–379.
- Savageau, M. A. (1970):** Biochemical systems analysis: III. Dynamic solutions using a power-law approximation. *Journal of Theoretical Biology* **26**(2), 215–226.
- Savageau, M. A. (1985):** A theory of alternative designs for biochemical control systems. *Biomedica Biochimica Acta* **44**(6), 875–880.
- Savageau, M. A.; Voit, E. O.; Irvine, D. H. (1987a):** Biochemical systems theory and metabolic control theory: 1. Fundamental similarities and differences. *Mathematical Biosciences* **86**(2), 127–145.
- Savageau, M. A.; Voit, E. O.; Irvine, D. H. (1987b):** Biochemical systems theory and metabolic control theory: 2. The role of summation and connectivity relationships. *Mathematical Biosciences* **86**(2), 147–169.
- Schaefer, U. (1999):** Automatisierte Probennahme zur Messung intrazellulärer Metabolitdynamiken. Dissertation, Universität Bonn.
- Schaefer, U.; Boos, W.; Takors, R.; Weuster-Botz (1999):** Automated Sampling Device for Monitoring Intracellular Metabolite Dynamics. *Analytical Biochemistry* **270**(1), 88–96.
- Schilling, C. H.; Edwards, J. S.; Palsson, B. O. (1999a):** Toward Metabolic Phenomics: Analysis of Genomic Data Using Flux Balances. *Biotechnology Progress* **15**(3), 288–295.
- Schilling, C. H.; Schuster, S.; Palsson, B. O.; Heinrich, R. (1999b):** Metabolic Pathway Analysis: Basic Concepts and Scientific Applications in the Post-Genomic Era. *Biotechnology Progress* **15**, 296–303.
- Schlegel, H. G. (1992):** *Allgemeine Mikrobiologie*. Thieme Verlag, Stuttgart, 7 Aufl.
- Schlesinger, A. (1979):** Terminology for model credibility. *Simulation* **32**(3), 103–104.
- Schlosser, P. M.; Riedy, G.; Baily, J. E. (1994):** Ethanol Production in Bakers's Yeast: Application of Experimental Techniques Perturbation Techniques for Model Development and Resultant Changes in Flux Control Analysis. *Biotechnology Progress* **10**(2), 141–154.
- Schmidt, K.; Carlsen, M.; Nielsen, J.; Villadsen, J. (1997):** Modeling Isotopomer Distributions in Biochemical Networks Using Isotopomer Mapping Matrices. *Biotechnology and Bioengineering* **55**(6), 831–840.
- Schmidt, P. P.; Travers, F.; Barman, T. (1995):** Transient and Equilibrium Kinetic Studies on Yeast 3-Phosphoglycerate Kinase. Evidence That an Intermediate Containing 1,3-Bisphosphoglycerate Accumulates in the Steady-State. *Biochemistry* **34**(3), 824–832.
- Schmit, J. C.; Artz, S.W.; Zalkin, H. (1970):** Chorismate Mutase-Prephenate Dehydrogenase. Evidence for Distinct Catalytic and Regulatory Sites. *The Journal of Biological Chemistry* **245**(16), 4019–4027.
- Schmitz, M. (2000):** Entwicklung analytischer Methoden zur Quantifizierung ausgewählter Metabolite in *Escherichia coli* K12. Diplomarbeit, Universität Bonn.
- Schomburg, I.; Chang, A.; Schomburg, D. (2002):** BRENDA: enzyme data and metabolic information. *Nucleic Acids Research* **30**(1), 47–49. <http://www.brenda.uni-koeln.de>.

- Schreyer, R.; Bock, A. (1980):** Phosphoglucose isomerase from *Escherichia coli* K10: purification, properties and formation under aerobic and anarobic conditions. *Archives of Microbiology* **127**(3), 289–298.
- Schuster, R.; Holzhütter, H. G.; Jacobasch, G. (1988):** Interrelations bwetween glycolysis and the hexose monophosphate shunt in erythrocytes as studied on the basis of a mathematical model. *BioSystems* **22**, 19–36.
- Schuster, S.; Dandekar, T.; Fell, D. A. (1999):** Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in Biotechnology* **17**(2), 53–60.
- Schuster, S.; Fell, D. A.; Dandekar, T. (2000):** A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology* **18**(3), 326–332.
- Schuster, S.; Pfeiffer, T.; Moldenhauer, F.; Koch, I.; Dandekar, T. (2002):** Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. *Bioinformatics* **18**(2), 351–361.
- Schwefel, H. P. (1995):** *Evolution and Optimum Seeking*. John Wiley & Sons Inc., New York, NY, USA.
- Segel, I. H. (1993):** *Enzyme Kinetics. Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. John Wiley & Sons, Inc, New York.
- Segel, L. A. (Hrsg.) (1988): *Mathematical models in molecular and cellular biology.*, Kap. A3, 649–727 University Press, Cambridge.
- Segel, L. A. (1999):** *Biological Kinetics*. Cambride University Press.
- Seigneur, C.; Stephanopoulos, G.; Carr Jr., R. W. (1982):** Dynamic sensitivity analysis of chemical reaction systems. A variational method. *Chemical Engineering Science* **37**(6), 845–853.
- Sel'kov, E. E. (1968):** Self-oscillations in glycolysis. 1. A simple kinetic model. *European Journal of Biochemistry* **4**(1), 79–86.
- Shanno, D. F. (1985):** An example of numerical nonconvergence of a variable-metric method. *Journal of Optimization Theory and Applications* **46**(1), 87–94.
- Siehs, C.; Oberbauer, R.; Mayer, G.; Lukas, A.; Mayer, B. (2002):** Discrete simulation of regulatory homo- and heterodimerization in the apoptosis effector phase. *Bioinformatics* **18**(1), 67–76.
- Orozco de Silva, A.; Fraenkel, D. G. (1979):** The 6-Phosphogluconate Dehydrogenase Reaction in *Escherichia coli*. *The Journal of Biological Chemistry* **254**(20), 10237–10242.
- Silverberg, M.; Dalziel, K. (1973):** Crystalline 6-Phospho-D-gluconate dehydrogenase from sheep liver. *European Journal of Biochemistry* **38**(2), 229–238.
- Smets, I.; Bernaerts, K.; Sun, J.; Marchal, K.; Vanderleyden, J.; van Impe, J. (2002):** Sensitivity Function-Based Model Reduction – A Bacterial Gene Expression Case Study. *Biotechnology and Bioengineering* **80**(2), 195–200.
- Smirnov, N. V. (1948):** Tables for estimating the goodness of fit of empirical distributions. *The Annals of Mathematical Statistics* **19**(2), 279–281.
- Smith, W. R.; Wronald, R. W.; Missen, R. W. (1991):** *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*. Wiley-Interscienc, New York, NY, USA.

- Smolen, P.; Baxter, D. A.; Byrne, J. H. (2000):** Modeling Transcriptional Control in Gene Networks: Methods, Recent Results, and Future Directions. *Bulletin of Mathematical Biology* **62**(2), 247–292.
- Sonnleitner, B.; Käppeli, O. (1986):** Growth of *Saccharomyces cerevisiae* is Controlled by its Limited Respiratory Capacity: Formulation and Verification of a Hypothesis. *Biotechnology and Bioengineering* **28**(6), 927–937.
- Spring, T. G.; Wold, F. (1971):** The Purification and Characterization of *Escherichia Coli* Enolase. *Journal of the Biological Chemistry* **52**(22), 6797–6802.
- Stelling, J.; Klamt, S.; Bettenbrock, K.; Schuster, S.; Gilles, E. D. (2002):** Metabolic network structure determines key aspects of functionality and regulation. *Nature* **420**(6912), 190–193.
- Stephanopoulos, G. N.; Aristidou, A. A.; Nielsen, J. (1998):** *Metabolic Engineering: Principles and Methodologies*. Academic Press, San Diego, London.
- Stock, J.B.; Waygood, E.B.; Meadow, N.D.; Postma, P.W.; Roseman, S. (1982):** Sugar Transport by the Bacterial Phosphotransferase System. The Glucose Receptors of the *Salmonella typhimurium* Phosphotransferase System. *The Journal of Biological Chemistry* **257**(28), 14543–14552.
- Stolarski, R. S.; Butler, D. M.; Rundel, R. D. (1978):** Uncertainty propagation in a stratospheric model 2. Monte Carlo analysis of imprecisions due to reaction rates. *Journal of Geophysical Research* **83**, 3074–3078.
- Sturm, T. F. (2002):** Operations Research. Skriptum zur Vorlesung im Herbsttrimester 2002 an der Universität der Bundeswehr München. <http://www.bwi.unibw-muenchen.de/home/sturm/skripten/opres2002.pdf>.
- Stute, W. (1997):** Nonparametric model checks for regression. *The Annals of Statistics* **25**(2), 613–641.
- Stute, W.; González Manteiga, W.; Presedo Quindimil, M. (1998):** Bootstrap Approximations in Model Checks for Regression. *Journal of the American Statistical Society* **93**(441), 141–149.
- Su, J. Q.; Wei, L. J. (1991):** A lack-of-fit test for the mean function in a generalized linear model. *Journal of the American Statistical Association* **86**(414), 420–426.
- Swartz, J. R. (1999):** *Escherichia coli* Recombinant DNA Technology. In: Neidhardt, F.C. (Hrsg.): *Escherichia Coli and Salmonella: Cellular and Molecular Biology*, Kap.108,1693–1712, ASM Press, Washington.
- Systeme des Lebens, Systembiologie (D) (2004):** <http://www.systembiologie.de>. Zugriffsdatum: 03/2004.
- Szyperski, T. (1998):** ¹³C-NMR, MS and metabolic flux balancing in biotechnology research. *Quarterly Reviews of Biophysics* **31**(1), 41–106.
- Takors, R. (1997):** Entwicklung und Einsatz einer Versuchsplanungstechnik zur experimentellen Makrokinetikbestimmung. Dissertation, Fachhochschule Aachen, Aachen.
- Termonia, Y.; Ross, J. (1981):** Oscillations and control features in glycolysis: Numerical analysis of a comprehensive model. *Proceedings of the National Academy of Sciences of the USA* **78**(5), 2952–2956.
- Teusink, B.; Passarge, J.; Reijenga, C. A.; Esgalhado, E.; van der Weijden, C. C.; Schepper, M.; Walsh, M. C.; Bakker, B. M.; van Dam, B.; van Dam, K.; Westerhoff, H. V.; Snoep, J. L. (2000):** Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry* **267**(17), 5313–5329.

- Theobald, U. (1995):** Untersuchungen zur Dynamik des Cabtree Effektes. Dissertation, Universität Stuttgart.
- Theobald, U.; Mailinger, W.; Baltes, M.; Rizzi, M.; Reuss, M. (1997a):** *In Vivo* Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: II. Mathematical Model. *Biotechnology and Bioengineering* **55**(4), 592–608.
- Theobald, U.; Mailinger, W.; Baltes, M.; Rizzi, M.; Reuss, M. (1997b):** *In Vivo* Analysis of Metabolic Dynamics in *Saccharomyces cerevisiae*: I. Experimental Observations. *Biotechnology and Bioengineering* **55**(2), 305–316.
- Theobald, U.; Mailinger, W.; Reuss, M.; Rizzi, M. (1993):** *In Vivo* Analysis of Glucose-Induced Fast Changes in Yeast Adenine Nucleotide Pool Applying a Rapid Sampling Technique. *Analytical Biochemistry* **214**(1), 31–37.
- Thomas, R. (1991):** Regulatory networks seen as asynchronous automata - a logical description. *Journal of Theoretical Biology* **153**(1), 1–23.
- Timmer, J.; Müller, T.; Melzer, W. (1998):** Numerical Methods to Determine Calcium Release Flux from Calcium Transients in Muscle Cells. *Biophysical Journal* **74**(4), 1694–1707.
- Tokarski, J. S.; Hopfinger, A. J. (1997):** Constructing Protein Models for Ligand-Receptor Binding Thermodynamic Simulations: An Application to a Set of Peptidomimetic Renin Inhibitors. *Journal of Chemical Information and Computer Science* **37**(4), 779–791.
- Tomita, M.; Hashimoto, K.; Takahashi, K.; Shimizu, T. S.; Matsuzaki, Y.; Miyoshi, F.; Saito, K.; Tanida, S.; Yugi, K.; Venter, J. C.; Hutchinson III, C. A. (1999):** E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**(1), 72–84.
- Torczon, V. (1991):** On the convergence of the multidirectional search algorithm. *SIAM Journal on Optimization* **1**(1), 123–145.
- Torralba, A. S.; Yu, K.; Shen, P. D.; Oefner, P. J.; Ross, J. (2003):** Experimental test of a method for determining causal connectivities of species in reactions. *Proceedings of the National Academy of Sciences of the USA* **100**(4), 1494–1498.
- Torres, J. C.; Babul, J. (1991):** An *in vitro* model showing different rates of substrate cycle for phosphofructokinases of *Escherichia coli* with different kinetic properties. *European Journal of Biochemistry* **200**(2), 471–476.
- Tribe, D. E.; Camakaris, H.; Pittard, J. (1976):** Constitutive and Repressible Enzymes of the Common Pathway of Aromatic Biosynthesis in *Escherichia coli* K-12: Regulation of Enzyme Synthesis at Different Growth Rates. *Journal of Bacteriology* **127**(3), 1085–1097.
- Turányi, T. (1990):** Sensitivity analysis of complex kinetic systems - Tools and applications. *Journal of Mathematical Chemistry* **5**(3), 203–248.
- Tweeddale, H.; Notley-McRobb, L.; Ferenci, T. (1998):** Effect of Slow Growth on Metabolism of *Escherichia coli*, as Revealed by Global Metabolite Pool (“Metabolome”) Analysis. *Journal of Bacteriology* **180**(19), 5109–5116.
- Uetz, P.; Giot, L.; Cagney, G.; Mansfield, T. A.; Judson, R. S.; Knight, J. R.; Lockshon, D.; Narayan, V.; Srinivasan, M.; Pochart, P.; Qureshi-Emili, A.; Li, Y.; Godwin, B.; Conover, D.; Kalbfleisch, T.; Vijajadamodar, G.; Yang, M.; Johnston, M.; Fields, S.; Rothberg, J. M. (2000):** A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**(6770), 623–627.

- Vajda, S.; Valko, P.; Turányi, T. (1985):** Principal Component Analysis of Kinetic Models. *International Journal of Chemical Kinetics* **17**, 55–81.
- Vallino, J. J.; Stephanopoulos, G. (2000):** Metabolic Flux Distributions in *Corynebacterium glutamicum* During Growth and Lysine Overproduction. *Biotechnology and Bioengineering* **67**(6), 872–885.
- Vance, W.; Arkin, A.; Ross, J. (2002):** Determination of causal connectivities of species in reaction networks. *Proceedings of the National Academy of Science of the USA* **99**(9), 5816–5821.
- Vanrolleghem, P. A.; Heijnen, J. J. (1998):** A Structured Approach for Selection Among Candidate Metabolic Network Models and Estimation of Unknown Stoichiometric Coefficients. *Biotechnology and Bioengineering* **58**(2-3), 133–138.
- Varma, A.; Boesch, B.; Palsson, B. (1993):** Biochemical Production Capabilities of *Escherichia coli*. *Biotechnology and Bioengineering* **42**(1), 59–73.
- Varma, A.; Palsson, B.O. (1994):** Metabolic flux analysis: basic concepts, scientific and practical use. *Bio/Technology* **12**, 994–998.
- Varner, J.; Ramkrishna, D. (1999a):** Metabolic Engineering from a Cybernetic Perspective: 1. Theoretical Preliminaries. *Biotechnology Progress* **15**(3), 407–425.
- Varner, J.; Ramkrishna, D. (1999b):** Metabolic Engineering from a Cybernetic Perspective: 2. Qualitative Investigation of Nodal Architectures and Their Response to Genetic Perturbation. *Biotechnology Progress* **15**(3), 426–438.
- Varner, J.; Ramkrishna, D. (1999c):** Metabolic Engineering from a Cybernetic Perspective: Aspartate Family of Amino Acids. *Metabolic Engineering* **1**, 88–116.
- Vaseghi, S.; Baumeister, A.; Rizzi, M.; Reuss, M. (1999):** *In vivo* Dynamics of the Pentose Phosphate Pathway in *Saccharomyces cerevisiae*. *Metabolic Engineering* **1**(2), 128–140.
- Vaseghi, S.; Macherhammer, F.; Zibek, S.; Reuss, M. (2001):** Signal Transduction Dynamics of the Protein Kinase-A/Phosphofructokinase-2 System in *Saccharomyces cerevisiae*. *Metabolic Engineering* **3**(2), 163–172.
- Venter, J. C (2001):** The sequence of the human genome. *Science* **291**(5507), 1304–1451.
- Veronese, F.M.; Boccù, E.; Fontana, A. (1976):** Isolation and Properties of 6-Phosphogluconate Dehydrogenase from *Escherichia coli*. Some Comparisons with the Thermophilic Enzyme from *Bacillus stearothermophilus*. *Biochemistry* **15**(18), 4026–4033.
- Versyck, K. J.; Claes, J. E.; Van Impe, J. F. (1997):** Practical Identification of Unstructured Growth Kinetics by Application of Optimal Experimental Design. *Biotechnology Progress* **13**(5), 524–531.
- Visser, D. (2001):** Measuring and Modeling *in vivo* kinetics of primary metabolism. Dissertation, Universität Delft.
- Visser, D.; van der Heijden, R.; Mauch, K.; Reuss, M.; Heijnen, S. (2000):** Tendency Modeling: A New Approach to Obtain Simplified Kinetic Models of Metabolism to *Saccharomyces cerevisiae*. *Metabolic Engineering* **2**(3), 252–275.
- Visser, D.; Heijnen, J. J. (2003):** Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic Engineering* **5**(3), 164–176.
- Visser, D.; van Zuylen, G. A.; van Dam, J. C.; Oudshoorn, A.; Eman, M. R.; Ras, C.; van Gulik, W. M.; Frank, J.; van Dedem, G. W. K.; Heijnen, J.J. (2002):** Rapid Sampling for Analysis of In

- Vivo Kinetics Using the BioScope: A System for Continuous-Pulse Experiments. *Biotechnology and Bioengineering* **79**(6), 674–681.
- Voit, E. O. (Hrsg.) (1991):** *Canonical nonlinear modeling. S-system approach to understanding complexity*. Kluwer Academic Publishers, Dordrecht.
- Wade, M.; Hertz, J. (2000):** Coarse-grained reverse engineering of genetic regulatory networks. *BioSystems* **55**(1-3), 129–136.
- Wagner, A. (2001):** How to reconstruct a large genetic network from n gene perturbation in fewer than n^2 easy steps. *Bioinformatics* **17**(11), 1183–1197.
- Walter, E. (1982):** *Identifiability of State Space Model: With Applications to Transformation Systems*. Springer-Verlag, New York, Berlin, Heidelberg.
- Walter, E.; Lecourtier, Y.; Happel, J. (1984):** On the Structural Output Distinguishability of Parametric Models, and its Relations with Structural Identifiability. *IEEE Transactions on Automatic Control* **AC-29**, 56–57.
- Walter, E.; Lecourtier, Y.; Raksanyi, A.; Happel, J. (1985):** On the distinguishability of parametric models with different structures. In: Eisenfeld, J.; De Lisi, C. (Hrsg.): *Mathematics and Computers in Biomedical Applications*, 145–160, Amsterdam, Elsevier.
- Walter, E.; Pronzato, L. (1996):** On the identifiability and distinguishability of nonlinear parametric models. *Mathematics and Computers in Simulation* **42**(2-3), 125–134.
- Watts, D. J.; Strogatz, S.H. (1998):** Collective dynamics of 'small-world' networks. *Nature* **393**(6684), 440–442.
- Westerhoff, H. V. (2001):** The Silicon Cell, Not Dead but Live! *Metabolic Engineering* **3**(3), 207–210.
- Westerhoff, H. V.; Bruggeman, F.; Hofmeyr, J. H. S.; Snoep, J. L. (2003):** Attractive models: how to make the silicon cell relevant and dynamic. *Comparative and Functional Genomics* **4**(1), 155–158.
- Westerhoff, H. V.; van Dam, K. (1987):** *Thermodynamics and Control of Biological Free-Energy Transduction*. Elsevier, Amsterdam.
- Westhead, D.; Clark, D. E.; Murray, C. W. (1997):** A comparison of heuristic search algorithms for molecular docking. *Journal of Computer-Aided Molecular Design* **11**(3), 209–228.
- Westwood, A. W.; Doelle, H. W. (1974):** Glucose-6-phosphate and 6-phosphogluconate dehydrogenases and their control mechanisms in *Escherichia coli* K12. *Microbios* **9**(35), 143–165.
- Weuster-Botz, D. (1997):** Sampling Tube Device for Monitoring Intracellular Metabolite Dynamics. *Analytical Biochemistry* **246**(2), 225–233.
- Whitner, R. B.; Balci, O. (1989):** Guidelines for selecting and using simulation model verification techniques. In: MacNair, E. A.; Musselman, K. J.; Heidelberger, P. (Hrsg.): *Proceedings of the 1989 Winter Simulation Conference*, 559–568, New York, NY, USA, ACM Press.
- Wiechert, W. (1990):** Interaktive Datenanalyse bei biotechnologischen Prozeßdaten. Dissertation, Universität Bonn.
- Wiechert, W. (1994):** Diskrete Analyse von Stoffwechselnetzwerken. In: Haubensack, F.; Sühnel, J. (Hrsg.): *Bioinformatik – Computereinsatz in den Biowissenschaften*, 100–111.
- Wiechert, W.; de Graaf, A. A. (1997):** Bidirectional Reaction Steps in Metabolic Networks: I. Modeling and Simulation of Carbon Isotope Labeling Experiments. *Biotechnology and Bioengineering* **55**(6), 101–117.

- Wiechert, W.; Siefke, C.; de Graaf, A. A.; Marx, A. (1997):** Bidirectional Reaction Steps in Metabolic Networks: II. Flux Estimation and Statistical Analysis. *Biotechnology and Bioengineering* **55**, 118–135.
- Wilson, J.; Hunt, T. (1994):** *Molecular Biology of the Cell: The Problems Book*. Garland Publishing, Inc., New York, USA, 2 Aufl.
- Wolf, J.; Heinrich, R. (2000):** Effect of cellular interaction on glycolytic oscillations in yeast: a theoretical investigation. *Biochemical Journal* **345**(2), 321–334.
- Wolf, J.; Sohn, H. Y.; Heinrich, R.; Kuriyama, H. (2001):** Mathematical analysis of a mechanism for autonomous metabolic oscillations in continuous culture of *Saccharomyces cerevisiae*. *FEBS Letters* **499**(3), 230–234.
- Wolkenhauer, O. (2001):** Systems biology: the reincarnation of systems theory applied in the biology? *Briefings in Bioinformatics* **2**(3), 258–270.
- Wood, D. O. (1986):** MIT model analysis program: What we learned about policy model review. In: Wilson, J. R.; Henriksen, J. O.; Roberts, S. D. (Hrsg.): *Proceedings of the 1986 Winter Simulation Conference*, 248–252, New York, NY, USA, ACM Press.
- Wright, B. E.; Albe, K. R. (1994):** Carbohydrate Metabolism in *Dictostelium discoideum*. I. Model Construction. *Journal of Theoretical Biology* **169**(3), 231–241.
- Wright, B. E.; Kelly, P. J. (1981):** Kinetic models of metabolism in intact cells, tissues and organisms. *Current Topics in Cellular Regulation* **19**, 103–158.
- Wright, R. D. (1972):** Validating dynamic models: An evaluation of tests of predictive power. In: *Proceedings of the 1972 Summer Computer Simulation Conference*, 1286–1296.
- Wu, C. F. J. (1986):** Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics* **14**(4), 1261–1350.
- Xia, X. Q.; Wise, M. J. (2003):** DiMSim: A Discrete-Event Simulator of Metabolic Networks. *Journal of Chemical Information and Computer Science* **43**(3), 1011–1019.
- Yang, X. S. (2003):** Turing pattern formation of catalytic reaction-diffusion systems in engineering applications. *Modeling and Simulation in Materials Science and Engineering* **11**(3), 321–329.
- Zangirolami, T. C.; Johansen, C. L.; Nielsen, J.; Joergensen, S. B. (1997):** Simulation of Penicillin Production in Fed-Batch Cultivations Using a Morphologically Structured Model. *Biotechnology and Bioengineering* **56**(6), 593–604.
- Zeigler, B. P. (1988):** Simplification of biochemical reaction systems. In: Segel, L. A. (Hrsg.): *Mathematical models in molecular and cellular biology*, University Press, Cambridge.
- Zeigler, B. P.; Praehofer, H.; Kim, G. K. (2000):** *Theory of Modelling and Simulation. Integrating Discrete Event and Continuous Complex Dynamic Systems*. Academic Press, San Diego, San Francisco, New York, 2 Aufl.

Anhang A Stoffwechselabschnitte

A.1 Glucose-Aufnahme

Bakterien können eine große Bandbreite von organischen Verbindungen als Nährstoffquellen nutzen, deren Translokation^{1,1} über die Cytoplasmamembran hinweg durch eine Reihe von Transportsystemen katalysiert wird. Abbildung A–1 zeigt ein allgemeines Schema der Phosphoenolpyruvat:Phosphotransferase Systeme (PTSs) (Postma et al. 1999), die für den Transport verschiedener Zucker verantwortlich sind und zugleich als Chemorezeptoren der Chemotaxis fungieren (Lux et al. 1995). Daneben übernehmen die Systeme regulatorische Funktionen. So hemmen z.B. Zucker, die über die PTSs aufgenommen werden, über die Regulationsvorgänge die Verwertung gleichzeitig vorhandener anderer Zucker (Katabolitrepression) (Ishizuka et al. 1993).

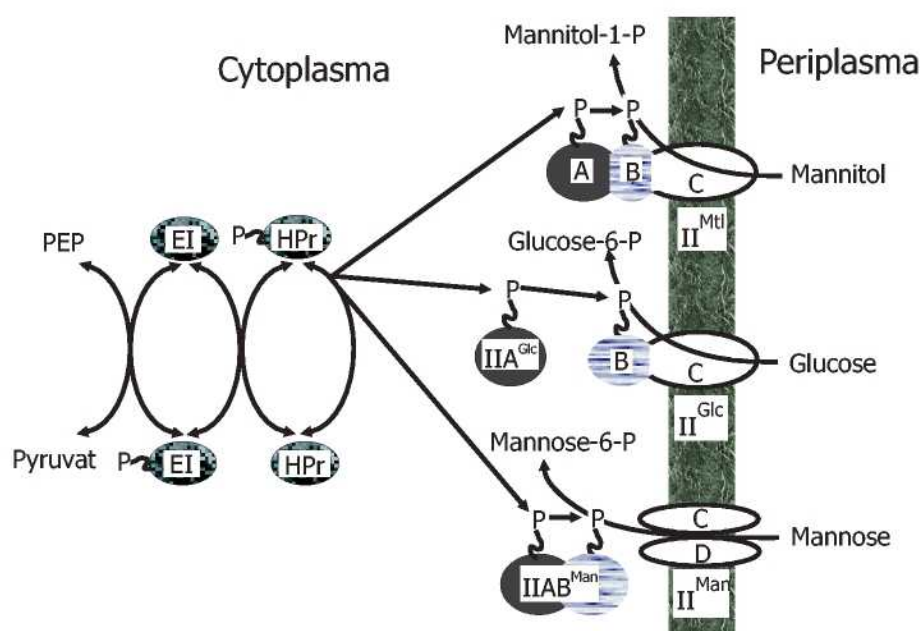
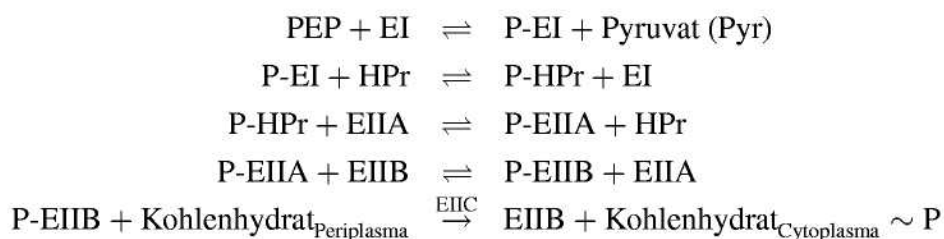


Abb. A–1 PEP:Phosphotransferase System, adaptiert aus (Postma et al. 1999, S. 1150) .

Während des durch die PTSs katalysierten Transportprozesses, der vor ca. 30 Jahren entdeckt wurde (Kundig et al. 1964), wird ein Zucker (Glucose (Gluc), Mannose (Man) oder Mannitol (Mtl)) durch PEP phosphoryliert, wobei Pyruvat entsteht. Der Prozeß läuft nach folgendem Reaktionsschema ab:



Im Gegensatz zu den löslichen Proteinen im Cytoplasma, Enzym I (EI) und HPr (Histidin-Protein), die allgemeine Bestandteile der PTSs darstellen, sind die Enzyme II für das jeweilige Kohlenhydrat

^{1,1}Translokation bezeichnet den aktiven Transport über die Membran, wobei das zu transportierende Molekül chemisch verändert wird.

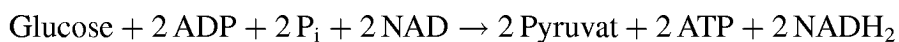
spezifisch. Sie können aus einem in der Membran gebundenen Protein mit drei Domänen (A, B und C), wie im Fall des Mannitol, bzw. aus zwei oder mehr Proteinen bestehen, von denen mindestens eines in der Membran gebunden (z.B. B und C) und eines löslich ist (EIIA, früher als Enzym III (EIII) bezeichnet, im Fall der Glucose). In jedem Fall wird die Phosphatgruppe von PEP über die obligaten Phosphatintermediate EI, HPr, EIIA, EIIB auf ein Kohlenhydrat übertragen. EIIC bildet als integrales Membranprotein den Translokationstunnel und besitzt vermutlich eine spezifische Substratbindungsregion.

In *E.coli* sind insgesamt 16 PTSs (EC 2.7.1.1.^{1,2}) bekannt, von denen 7 Glucose mit unterschiedlichen Geschwindigkeiten transportieren können (Lengeler 1993). Die ausgeprägte Affinität von *E.coli* für Glucose ($K_m=3-10 \mu\text{M}$ (Misset et al. 1983; Stock et al. 1982)) hängt mit der hohen Substratspezifität des Enzyms II^{Glc} zusammen. Dieses ist ein wesentliches Enzym für die Regulierung der Katabolitrepresion (Postma et al. 1993). In seiner phosphorylierten Form aktiviert es die Adenylat-Cyclase, die dafür sorgt, daß in einer hungernden Zelle eine hohe Konzentration von cyclischem Adenosinmonophosphat (cAMP) vorhanden ist. Dieses bildet einen aktiven Komplex mit dem cAMP-Rezeptorprotein (CRP), der an alle Promotoren mit einer spezifischen CRP-Bindungssequenz (z.B. dem *lac*-Operon) bindet. Bei Anwesenheit von weiteren Induktoren (wie z.B. Lactose) für die betreffenden Operons ist damit die Transkription alternativer Transportsysteme möglich. Wenn PTS-Zucker schnell in die Zelle transportiert werden, sinkt die Konzentration des phosphorylierten II^{Glc} und damit auch der cAMP Spiegel, so daß die CRP-abhängigen Promotoren reprimiert bleiben und vorwiegend das Transportsystem für den im Überschuß vorliegenden Zucker exprimiert wird.

Obwohl der Transport und die Phosphorylierung der Glucose gewöhnlich durch die PTSs katalysiert wird, kann Glucose auch über Nicht-PTS-Transporter in die Zelle gelangen. Für die Phosphorylierung der intrazellulären Glucose, die u.a. auch aus Lactose gebildet werden kann, besitzen Enterobakterien ein weiteres Enzym, die Glucokinase. Kofaktor dieser Reaktion ist ATP.

A.2 Glykolyse

Die Glykolyse, auch bekannt als Embden-Meyerhof-Parnas-Weg (EMP-Weg), ist neben dem Pentose-Phosphat-Weg (PPP), s. Abschnitt A.3, der wichtigste Reaktionsweg zur Verstoffwechselung von Zuckern (70% der Glucose werden unter aeroben Bedingungen in der Glykolyse bzw. 30% im PPP verstoffwechselt (Schlegel 1992)). Unter der Voraussetzung, daß die gesamte Menge der aufgenommenen Glucose über die Glykolyse abgebaut wird, läßt sich die folgende Bilanzgleichung aufstellen:



Die Glykolyse liefert damit chemische Energie in Form von ATP und in Form von NADH_2 , das in der Atmungskette unter ATP-Bildung durch Sauerstoff oxidiert werden kann. Ein detaillierteres Bild der Glykolyse ist in Schema A-2 gezeigt (Schlegel 1992).

Die Glucose wird während der Translokation über die Cytoplasmamembran zu Glucose-6-phosphat (G6P) phosphoryliert, s. Abschnitt A.1. Die anschließende Isomerisierung von Glucose-6-phosphat (G6P) zu Fructose-6-phosphat (F6P) wird durch die Phosphoglucoseisomerase katalysiert. Daraufhin wird F6P über die 6-Phosphofructokinase mit ATP in 1-Stellung phosphoryliert. Das wesentliche Enzym dieser Reaktion, Pfk-1 kodiert durch das *pfkA*-Gen, ist ein allosterisches Enzym, das von Nucleosiddiphosphaten aktiviert bzw. durch PEP inhibiert wird, und eine wichtige Funktion für die metabolische

^{1,2}EC bezeichnet die Enzymklassifizierungsnummer. Die erste Ziffer steht für den Klassennamen (2: Transferase), die zweite für die Unterklasse (7: Phosphotransferase), die dritte Ziffer (1) gibt an, daß eine Hydroxygruppe als Akzeptor der Phosphatgruppe dient, die vierte Ziffer (1), daß Glucose als Akzeptormolekül dient.

Regulation der Glykolyse übernimmt. Das entstandene Fructose-1,6-bisphosphat (FBP) wird durch die Fructose-1,6-bisphosphat-Aldolase in zwei C_3 -Fragmente, Dihydroxyacetonphosphat (DHAP) und Glyceraldehyd-3-phosphat (GAP) gespalten. Ihre reversible Isomerisierung wird durch die Triosephosphat-Isomerase katalysiert. Das Gleichgewicht der Reaktion liegt zwar auf der Seite des DHAP, die Umwandlung des Ketose- in das Aldose-Isomer geschieht jedoch aufgrund der hohen katalytischen Aktivität der Isomerase sehr schnell, so daß die C_3 -Fragmente über GAP weiter verstoffwechselt werden können. Ein Teil der bei der Oxidation von GAP zu 3-Phosphoglycerat (3PG) frei werdenden Energie

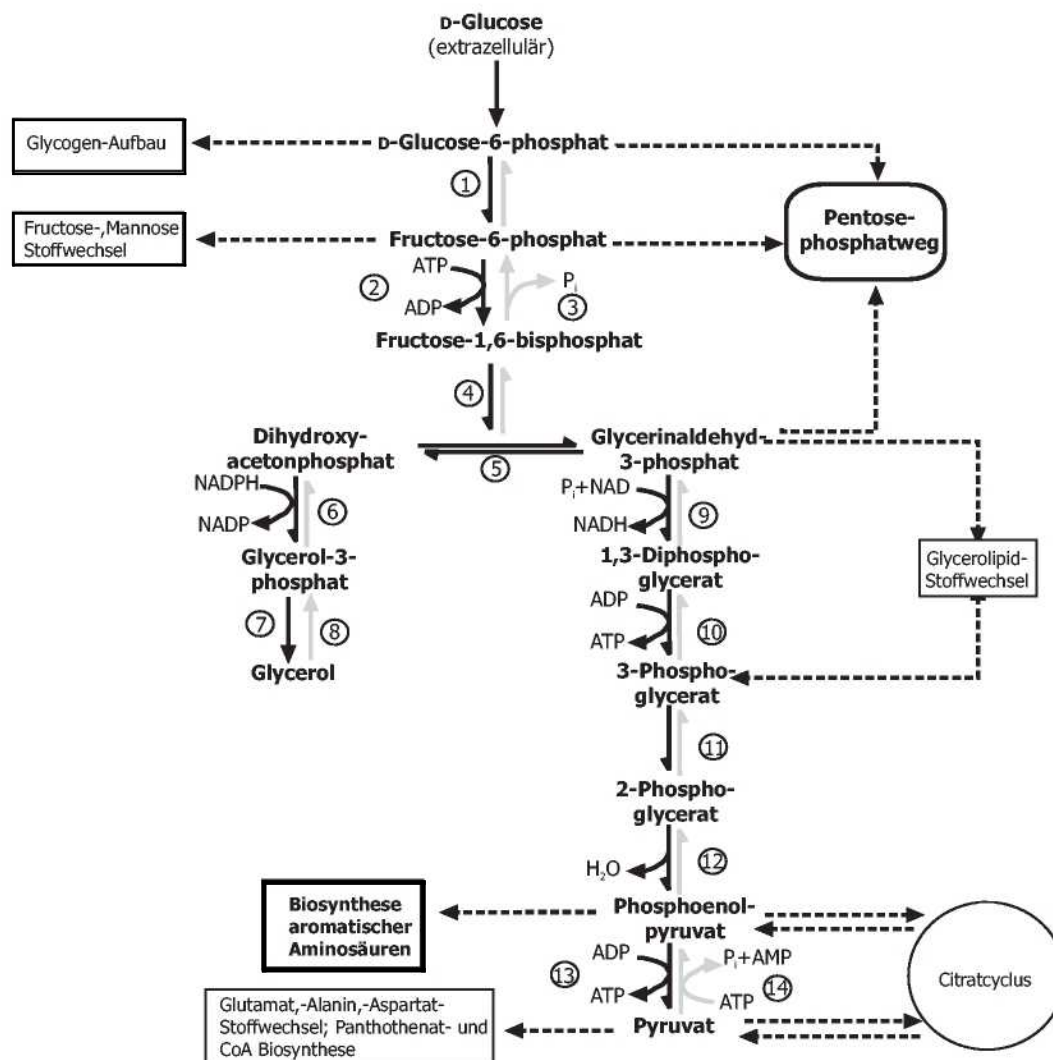


Abb. A-2 Reaktionen der Glykolyse (schwarze Pfeile) und der Gluconeogenese (graue Pfeile), s. (Schlegel 1992); 1: Phosphoglucisomerase (EC 5.2.1.9, Gensymbol: *pgi*); 2: 6-Phosphofructokinase (EC 2.7.1.11, *pfkA*, *pfkB*); 3: Fructose-1,6-bisphosphatase (EC 3.1.3.11, *fbp*); 4: Fructosebisphosphat-Aldolase (EC 4.1.2.13, *fbaA*); 5: Triosephosphat-Isomerase (EC 5.3.1.1, *tpiA*); 6: Glycerol-1-phosphatase (EC 3.1.3.21, *gpp*); 7: Glycerol-3-phosphat-Dehydrogenase (EC 1.1.99.5, *glpB*, *glpC*, *glpD*); 8: Glycerolkinase (EC 2.7.1.30, *glpK*); 9: Glyceraldehyd-3-phosphat-Dehydrogenase (EC 1.2.1.12, *gapA*); 10: Phosphoglyceratkinase (EC 2.7.2.3, *pgk*); 11: Phosphoglyceratmutase (EC 5.4.2.1, *gpm*); 12: Enolase (EC 4.2.1.11, *eno*); 13: Pyruvat-Kinase (EC 2.7.1.40, *pykA*, *pykF*); 14: Phosphoenolpyruvat-Synthase (EC 2.7.9.2, *pps*).

wird durch Phosphorylierung von ADP konserviert. Dazu wird zunächst die Aldehydgruppe von GAP an eine Hydrosulfidgruppe der Glyceraldehyd-Dehydrogenase gebunden, dann ein Proton abgespalten und auf NAD übertragen. Über Phosphorolyse wird die Thiogruppe gegen eine Phosphatgruppe

ausgetauscht, 1,3-Diphosphoglycerat (DPG) entsteht, das die energiereiche Phosphatbindung auf ADP überträgt, wobei 3-Phosphoglycerat (3PG) und ATP gebildet werden. Die Oxidation von GAP läuft nur in der Anwesenheit von Orthophosphat und ADP ab, ansonsten kommt die Glykolyse zum Stillstand (Pasteur-Effekt). Durch die Phosphoglyceromutase wird 3PG zu 2-Phosphoglycerat (2PG) umgewandelt, woraus unter Wasserabspaltung PEP entsteht. PEP ist ein Enolester, dessen energiereiche Phosphatgruppe auf ADP übertragen wird. Das entstandene Pyruvat (Pyr) ist die Vorstufe weiterer Synthese- und Abbauprozesse, wie dem Citratzyklus unter aeroben Bedingungen oder der Lactat-, Acetat-, Succinat- oder Formiat-Gärung unter anaeroben Bedingungen (Schlegel 1992). Der Energiegewinn ist im letztem Fall schlechter, d.h. höchstens 4 mol ATP je mol Glucose können im Gegensatz zu den 26 mol ATP beim Durchlaufen des Citratzyklus synthetisiert werden.

Durch hohe Glucose-Konzentrationen mit ausreichender Sauerstoffversorgung in der Umgebung kann ein sehr hoher Stofffluß durch die Glykolyse erzielt werden. Es kommt zu einer Art „Glucose-Überschußmetabolismus“, so daß die Endprodukte der Glykolyse akkumulieren, in Acetat umgewandelt und von der Zelle ausgeschieden werden. Dies stellt eine partielle aerobe Gärung dar, die als Crabtree-Effekt bezeichnet wird.

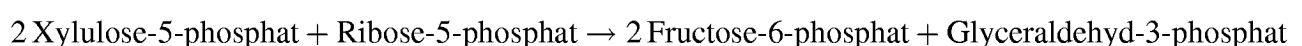
Die Gluconeogenese ist die enzymgesteuerte Neubildung von Kohlenhydraten durch die Umkehr der Glykolyse. Die Gluconeogenese ist in Mikroorganismen aktiv, die in einem Acetat-, Succinat-, Lactat- oder Glycerol-haltigen Medium wachsen. Sie erfolgt über die gleichen Intermediate wie Glykolyse, benötigt aber Energie in Form von ATP. Die durch die 6-Phosphofructokinase und Pyruvat-Kinase katalysierten Reaktionen werden durch Enzymreaktionen umgangen, die in Richtung der Glucose-Synthese exergonisch sind. So wird die Pyruvat-Kinase-Reaktion durch die Reaktion über die Phosphoenolpyruvat-Synthase umgangen, die Pyruvat zu PEP phosphoryliert. Die Reaktion von Fructose-1,6-bisphosphat zu Fructose-6-phosphat wird durch die Fructose-1,6-bisphosphatase katalysiert und umgeht somit die 6-Phosphofructokinase-Reaktion.

A.3 Pentose-Phosphat-Weg

Im zellulären Stoffwechsel erfüllt der Pentose-Phosphat-Weg (PPP) mehrere Aufgaben, s. Abbildung A-3. Er dient der Synthese von Reduktionsäquivalenten für die Biosynthesewege in der Form von NADPH oder stellt Ribose-5-phosphat bereit, das zum Aufbau der Nucleotide benötigt wird.

Im oxidativen Zweig wird Glucose-6-phosphat durch die Glucose-6-phosphat-Dehydrogenase unter Bildung des 6-Phosphogluconolactons oxidiert, das spontan oder durch die Gluconolactonase katalysiert zu 6-Phosphogluconat (6PG) hydrolysiert. 6PG wird durch die 6-Phosphogluconat-Dehydrogenase zu 3-Keto-6-phosphogluconat dehydrogeniert und zu Ribulose-5-phosphat decarboxyliert. Damit ist der oxidative Teil des PPP abgeschlossen. Das Verhältnis zwischen dem hier gebildeten NADPH zu dem entstandenen Pentose-5-phosphat beträgt 2:1.

Die folgenden Reaktionen sind Umwandlungsprozesse zwischen Pentose- und Hexosephosphaten. Gibt es in der Zelle einen erhöhten Bedarf an NADPH, wird der gebildete Überschuß an Pentose-5-phosphaten über die Umwandlungsreaktionen abgebaut. Dazu katalysiert die Transaldolase den Transfer einer Dihydroxyaceton-Gruppe von Sedoheptulose-7-phosphat (S7P) auf GAP unter Bildung von Erythrose-4-phosphat und F6P. Die Transketolase überträgt eine Glycolaldehyd-Gruppe von Xylulose-5-phosphat auf Ribose-5-phosphat oder Erythrose-4-phosphat, wobei S7P und GAP entstehen. Beide Enzyme zusammen katalysieren die Umwandlung von Pentose-5-phosphaten in F6P und GAP, die der Glykolyse zugeführt werden können:



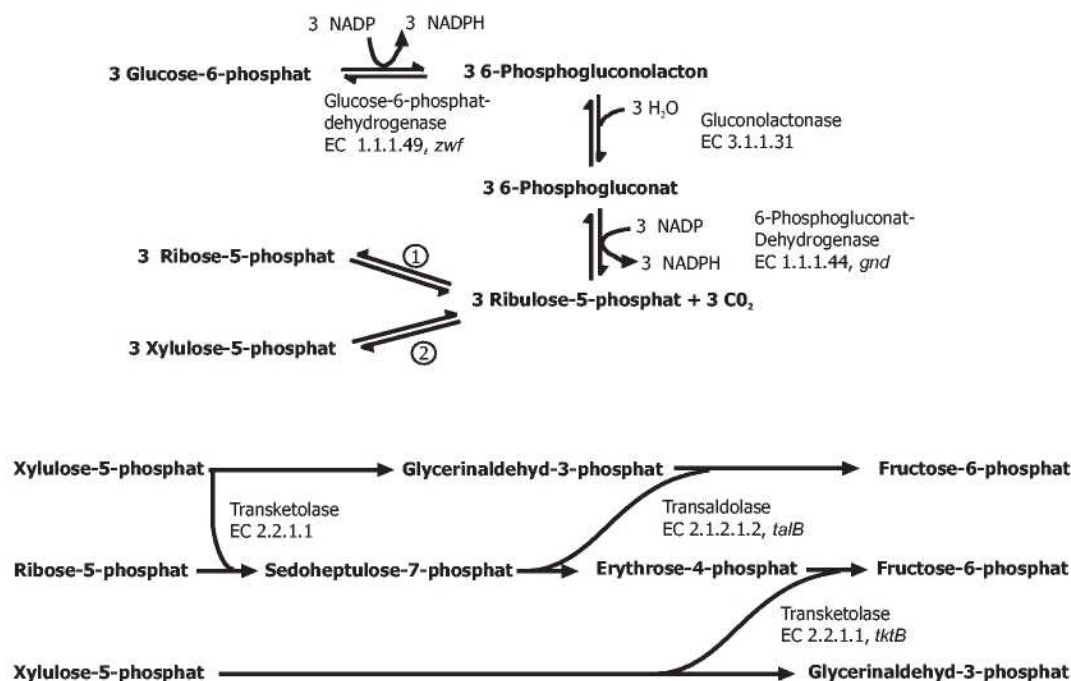


Abb. A-3 Schema des Pentose-Phosphat-Weges; 1: Ribose-5-phosphat-Isomerase (EC 5.3.1.6, *rpiA*); 2: Ribulose-5-phosphat-3-Epimerase (EC 5.1.3.1, *rpe*)

Alle Reaktionen sind reversibel, so daß umgekehrt Pentose-5-phosphate aus F6P und GAP gebildet werden können, wenn die Zelle Vorläuferstoffe für die Nukleotid- oder Aromatenbiosynthese benötigt, aber keinen erhöhten Bedarf an NADPH hat. Da auch die Reaktionen der Glykolyse reversibel sind, d.h. die Isomersierung von F6P zu G6P und die Kondensation von zwei Triosephosphaten zu einem Hexose-6-phosphat möglich ist, kann sich ein Zyklus zwischen Glykolyse und PPP ergeben, bei dessen einmaligem Umlauf aus drei Molekülen G6P zwei Moleküle F6P, ein Molekül GAP, drei CO_2 -Moleküle und drei Moleküle NADPH entstehen.

A.4 Synthese aromatischer Aminosäuren

Die Synthese der aromatischen Aminosäuren in *Escherichia coli* ist in Abbildung A-4 dargestellt. Der zunächst für alle Aminosäuren gleiche Biosyntheseweg (Shikimat-Weg) spaltet sich ausgehend vom Chorismat in die Synthesewege für L-Phenylalanin, L-Tryptophan und L-Tyrosin auf. Vier weitere hier nicht gezeigte Biosynthesewege gehen vom Chorismat aus zur Folsäure, Ubichonon, Menachinon und Enterocholin (Gibson 1964).

Drei isofunktionelle Metalloenzyme, die DAHP-Synthasen (3-Deoxy-D-arabino-heptulosonat-7-phosphat-Synthasen) kondensieren Erythrose-4-phosphat (E4P), einen Vorläuferstoff aus dem PentosePhosphat-Weg, und PEP aus der Glykolyse zu DAHP. Neben strukturellen Unterschieden können die Isoenzyme funktionell aufgrund ihrer jeweiligen Inhibierungen durch die Endprodukte der Biosynthesewege unterschieden werden. Das L-Phenylalanin-spezifische Isoenzym, das durch das Gen *aroG* kodiert wird, wird durch L-Phenylalanin (Phe) bis zu 95% inhibiert. Ebenso wird die Aktivität der durch das Gen *aroF* kodierten, L-Tyrosin spezifischen DAHP-Synthase um bis zu 95% durch L-Tyrosin (Tyr) gesenkt. Das durch L-Tryptophan (Trp) bis zu 60% inhibierte Isoenzym ist Trp-spezifisch und wird durch das Gen *aroH* kodiert (Gibson und Pittard 1968). Zudem ist bekannt, daß die DAHP-Synthase(Tyr) weiterhin durch hohe Phe-Konzentrationen und die DAHP-Synthase(Phe) auch durch Trp inhibiert wird (Brown und Sommerville 1971). Neben der Inhibierung des ersten Syntheseschritts durch die

Endprodukte der Biosynthese („Feedback“-Inhibierung) wird die Expressionsrate der Enzyme auf der genetischen Ebene durch die Aminosäuren reguliert. Die Synthese der DAHP-Synthase(Tyr) kann durch die Aktivierung des *tyrR*-Regulons in Anwesenheit von Tyr oder auch sehr hohen Phe-Konzentrationen reprimiert werden. Gleiches gilt für die DAHP-Synthase(Phe) (Pittard 1999). Bei Wachstum auf einem Minimalmedium übernimmt die DAHP-Synthase(Phe) mit bis zu 80% der kompletten Syntheseaktivität den Hauptanteil an der Katalyse. Wenn aromatische Aminosäuren das Wachstum limitieren, ist die DAHP-Synthase(Tyr) das wichtigste Enzym für die Synthese (Tribe et al. 1976).

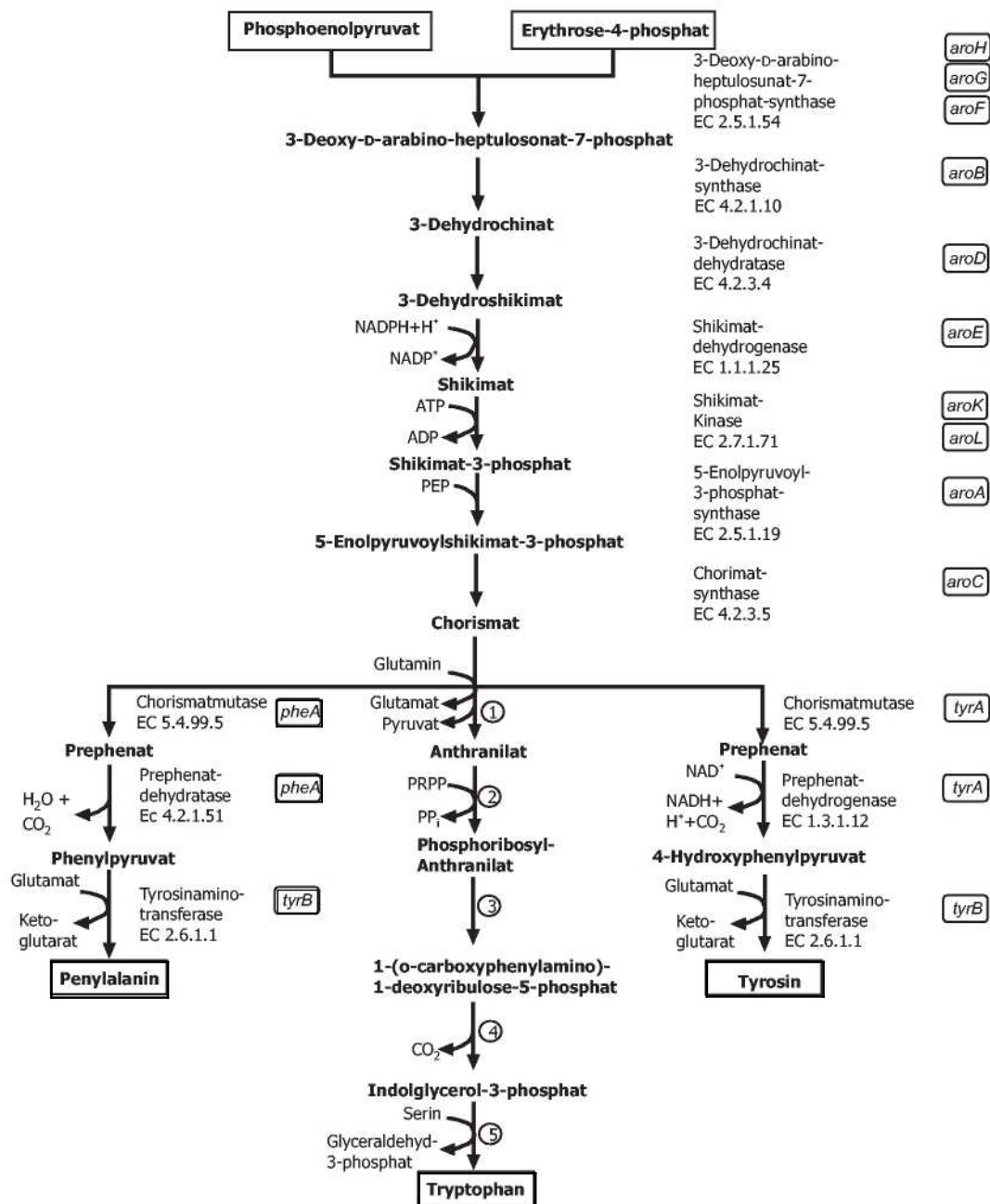


Abb. A-4 Schema der Synthese aromatischer Aminosäuren; 1: Anthranilat-Synthase (EC 4.1.3.27, *trpE*, *trpD*); 2: Anthranilat-Phosphoribosyl-Transferase (EC 2.4.2.18, *trpD*); 3: Phosphoribosyl-Anthranilat-Isomerase (EC 5.3.1.24, *trpC*); 4: Indolglycerol-3-phosphat-Synthase (EC 4.1.1.48, *trpC*); 5: Tryptophan Synthase (EC 4.2.1.20, *trpA*, *trpB*).

Die 3-Dehydrochinat-Synthase (DHQ-Synthase) wandelt DAHP zu DHQ um. Obwohl die Expression der DHQ-Synthase, wie auch aller weiterer Enzyme, soweit nicht anders angegeben, konstitutiv ist,

können in diesem Schritt Limitierungen der Aromatenbiosynthese aufgrund einer Substratinhibierung des Enzyms durch DAHP auftreten (Ogino et al. 1982).

Die DHQ-Dehydratase katalysiert die Umsetzung von DHQ zu 3-Dehydroshikimat (DHS) und führt die erste Doppelbindung des aromatischen Rings ein. DHS wird unter NADPH-Verbrauch von der Shikimat-Dehydrogenase zu Shikimat (Shik) reduziert. Die darauffolgende Reaktion von Shik zu Shikimat-3-phosphat (S3P) wird durch zwei Isoenzyme Shikimat-Kinase I und II katalysiert, wobei ATP als Kofaktor für die Phosphorylierung benötigt wird. Die Expressionsrate der durch das Gen *aroL* kodierten Shikimat-Kinase II wird über die Expression des *tyrR*-Regulons bei einem Überschuß von Tyr oder Trp reprimiert, so daß die Enzymmenge nur bis zu einem Zehntel der normal vorhandenen Menge betragen kann (Ely und Pittard 1979). Die Shikimat-Kinase I wird genetisch nicht reguliert und hat mit einem K_m -Wert von 5 mM eine weitaus geringere Substrataffinität als die Kinase II ($K_m=200 \mu\text{M}$) (Pittard 1999), so daß diskutiert werden kann, ob dieses Enzym unter Umständen noch andere Stoffwechselfunktionen übernimmt.

Unter Verbrauch eines weiteren PEP-Moleküls entsteht in der durch die 5-Enolpyruvoyl-3-phosphat-Synthase (EPSP) katalysierten Reaktion EPSP aus S3P. Die Chorismat-Synthase führt die zweite Doppelbindung des aromatischen Rings in EPSP ein und bildet Chorismat, das den Ausgangspunkt der sich nun aufteilenden Synthesewege der aromatischen Aminosäuren darstellt.

Bifunktionelle Enzyme katalysieren die ersten beiden Reaktionen des Phe- und des Tyr-Weges. Die erste Reaktion, in der Prephenat durch die Chorismat-Mutase gebildet wird, ist für beide Wege gleich. Anschließend wird das entstandene Prephenat durch die zweite Funktionseinheit der Enzyme, die Prephenat-Dehydratase im Phe-Weg bzw. die Prephenat-Dehydrogenase im Tyr-Weg zu Phenylpyruvat oder 4-Hydroxyphenylpyruvat umgesetzt. Diese Schritte werden stark reguliert. Phe hemmt die Chorismat-Mutase / Prephenat-Dehydratase (kodiert durch *pheA*), Tyr inhibiert die Chorismat-Mutase / Prephenat-Dehydrogenase, die durch *tyrA* kodiert wird (Pittard 1999). Phe kann damit die erste Aktivität des bifunktionellen Enzyms bis zu 90%, die zweite Aktivität bis maximal zu 55% hemmen (Schmit et al. 1970; Dopheide et al. 1972). Auf der Ebene der Transkription wird die Expression von *pheA* durch Attenuation reguliert (Borg-Olivier et al. 1987). Die letzte Reaktion zu Phe bzw. Tyr ist die Aminierung der α -Ketosäuren mit Glutamat als Aminogruppen-Donor, die im Wesentlichen durch die Asparat- oder die Tyrosin-Aminotransferasen katalysiert wird.

Der Tryptophan-Weg stellt keinen Gegenstand der Untersuchungen dieser Arbeit dar, weshalb an dieser Stelle auf eine Beschreibung verzichtet und auf andere Zusammenfassungen des Biosyntheseweges verwiesen wird (Pittard 1999).

Anhang B Verfahrenstechnische Grundlagen

Bioreaktoren werden grundsätzlich bezüglich des Reaktortyps (Rührkessel, Säulenreaktor, Strömungsröhr) und der Betriebsführung eingeteilt (Dunn et al. 2003). Über die Betriebsführung werden im Wesentlichen die Zeitprofile der Substrat- und Produktkonzentrationen bestimmt, die sich wiederum häufig auf die Raten der biologischer Reaktionen (Wachstum, Substratverbrauch, Produktbildung) auswirken. Für alle Betriebsarten gilt die allgemeine Bilanzgleichung:

$$\underbrace{\frac{d(V_R \cdot c_R)}{dt}}_{\text{Akkumulation}} = \underbrace{\dot{F}_{in} \cdot c_{in} - \dot{F}_{out} \cdot c_R}_{\text{Konvektion}} \pm \underbrace{r_i V_R}_{\text{Reaktion}}, \quad (\text{B-1})$$

wobei $r_i \cdot V_R$ die reaktorbezogene Reaktionsrate (Produktbildungsrate π oder Substratverbrauchsrate σ), $\dot{F}_{in}$ der Volumenstrom in den Reaktor bzw. $\dot{F}_{out}$ der Volumenstrom aus dem Reaktor, V_R das Arbeitsvolumen des Reaktors ist, und c_{in} bzw. c_R die Konzentrationen der Produkte, Substrate und der Biomasse im Zulauf bzw. im Reaktor darstellen.

Im *batch*- oder *satzweisen* Verfahren werden alle Substrate vorgelegt und während der Fermentation nur Sauerstoff bei aeroben Prozessen oder Korrekturmittel (Lauge, Antischaummittel) zugegeben, d.h. $V_R = \text{konst.}$ und $\dot{F}_{in} = \dot{F}_{out} = 0$. Die Massenbilanzgleichung vereinfacht sich also zu:

$$\frac{dc_R}{dt} = \pm r_i \quad (\text{B-2})$$

Bei einer logarithmischen Auftragung der Zellzahl bzw. der Biomasse über der Zeit wird bei optimalen Bedingungen und nur einer Substratquelle eine für diesen Prozeß typische Wachstumskurve beobachtet, s. Abbildung B-1. Die Adaption- oder lag-Phase umfaßt das Zeitintervall zwischen Einbringen der

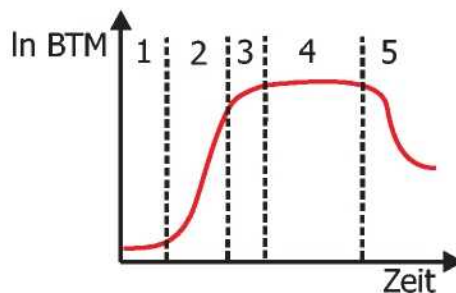


Abb. B-1 Wachstumskurve: Adaption- oder lag-Phase (1); exponentielles Wachstum (2); Übergangsphase (3); stationäre Phase (4) und Absterbephase (5).

Vorkultur in den Bioreaktor (Animpfen) und dem ersten merklichen Wachstum, in dem sich die Mikroorganismen an die neue Umgebung anpassen. Während der exponentiellen Phase folgt das Zellwachstum dem allgemeinen Gesetz für ein freies oder geometrisches Wachstum $d BTM / dt = \mu_{max} BTM$, wobei μ_{max} die spezifische (maximale) Wachstumsgeschwindigkeit bezeichnet. Durch die Limitierung eines essentiellen Substrates oder Ausscheidung eines inhibierenden bzw. toxischen Produktes wird ein weiteres exponentielles Wachstum verhindert, d.h. die exponentielle Phase mündet über eine Übergangsphase in die stationäre Phase ein. In dieser Phase hat die Kultur ihre maximale Zellkonzentration erreicht, in der sich Zellwachstum und Absterben die Waage halten. Die Dauer der stationären Phase ist von der Menge der in den Zellen beinhaltenen Reservestoffen abhängig. Es schließt sich die Absterbephase an, in der die Autolyse der Zellen überwiegt.

Während die *Batch*-Kultur durch sich ständig ändernde Umgebungsbedingungen gekennzeichnet ist, an die sich die Mikroorganismen entsprechend anpassen müssen, werden bei der *kontinuierlichen* Kultivierung der Zellen die Prozeßparameter konstant gehalten. Dabei wird dem Reaktor nach der Adaptionsphase sterile Nährlösung kontinuierlich zugegeben und gleichzeitig die gleiche Menge umgesetzte Nährlösung mit Mikroorganismen dem System entnommen. Somit gilt:

$$V_R = \text{konst.} \quad \text{und} \quad \dot{F}_{in} = \dot{F}_{out} = \text{konst.}$$

Unter Verwendung der Durchflußrate D , bzw. der mittleren Verweilzeit $\bar{\tau}$, die folgendermaßen definiert wird:

$$D \stackrel{\text{def}}{=} \frac{1}{\bar{\tau}} = \frac{\dot{F}}{V_R},$$

kann die vereinfachte Bilanzgleichung für kontinuierliche Verfahren abgeleitet werden:

$$\frac{dc_R}{dt} = D(c_{in} - c_R) \pm r_i. \quad (\text{B-3})$$

Für den stationären Fall, d.h. im Fließgleichgewicht, ist $dc_R/dt = 0$, damit lautet die Bilanzgleichung: $D(c_{in} - c_R) = r_i$. Ist im Fall eines mikrobiellen Wachstumsprozesses $BTM_{in} = 0$, so erhält man unter Verwendung des obigen Gesetz für ein geometrisches Wachstum der Zellen:

$$D = \mu, \quad (\text{B-4})$$

d.h. die spezifische Wachstumsgeschwindigkeit der Zellen hängt allein von der Verweilzeit der Biomasse im Reaktor ab.

Bei *Fermentationen im Zulaufverfahren* (engl: fed-batch) werden im Laufe des Prozesses ein oder mehrere Substanzen dem Reaktor zugeführt. Die Zufuhr erfolgt nach einem vorgegebenen Dosierungsprofil (Steuerung) oder wird unter Einbeziehung direkter bzw. indirekter Meßgrößen geregelt. Es wird keine Fermentationslösung abgezogen. Somit gilt: $c_R = f(t)$, $V_R = f(t)$, $\dot{F}_{in} = f(t)$ und $\dot{F}_{out} = 0$. Somit ergibt sich die Bilanzgleichung für das Fed-Batch-Verfahren:

$$\frac{d(V_R \cdot c_R)}{dt} = \dot{F}_{in} \cdot c_{in} \pm r_i \cdot V_R. \quad (\text{B-5})$$

Die Wahl des Betriebsverfahrens für einen Prozeß wird maßgeblich von der Reaktionskinetik bestimmt. Beispielsweise verzögern katabolitrepriemierende Prozesse das Wachstum in einer satzweisen Fermentation, während die Verzögerung des Wachstums bei einer Fed-Batch-Fermentation aufgrund der geringen Substratkonzentration im Reaktor verhindert werden kann.

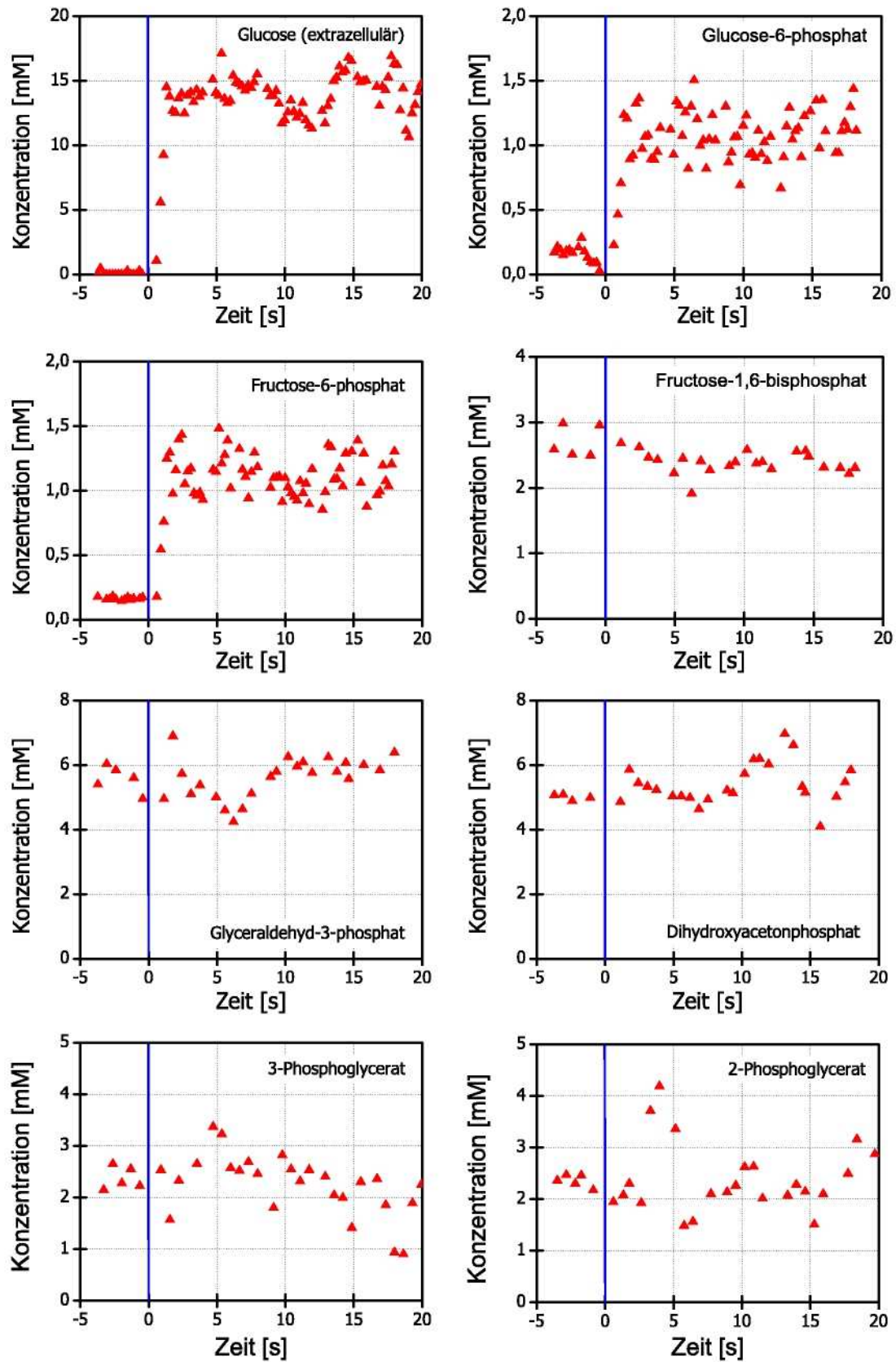


Abb. C-2 Trajektorien von glykolytischen Intermediaten aus Experiment WT1; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

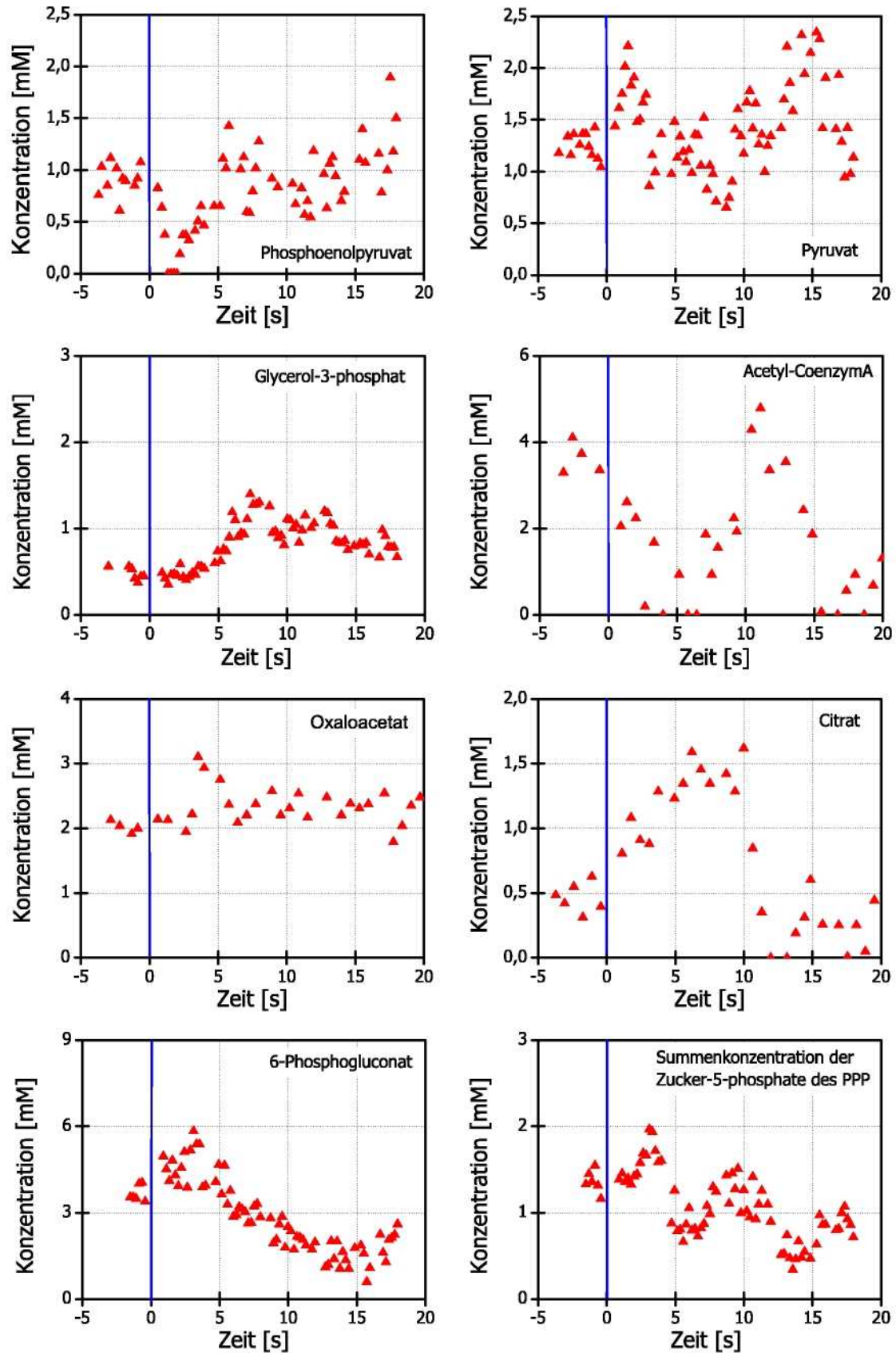


Abb. C-3 Trajektorien von Intermediaten aus der Glykolyse, dem Citronensäurezyklus und dem Pentose-Phosphat-Weg aus Experiment WT1; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

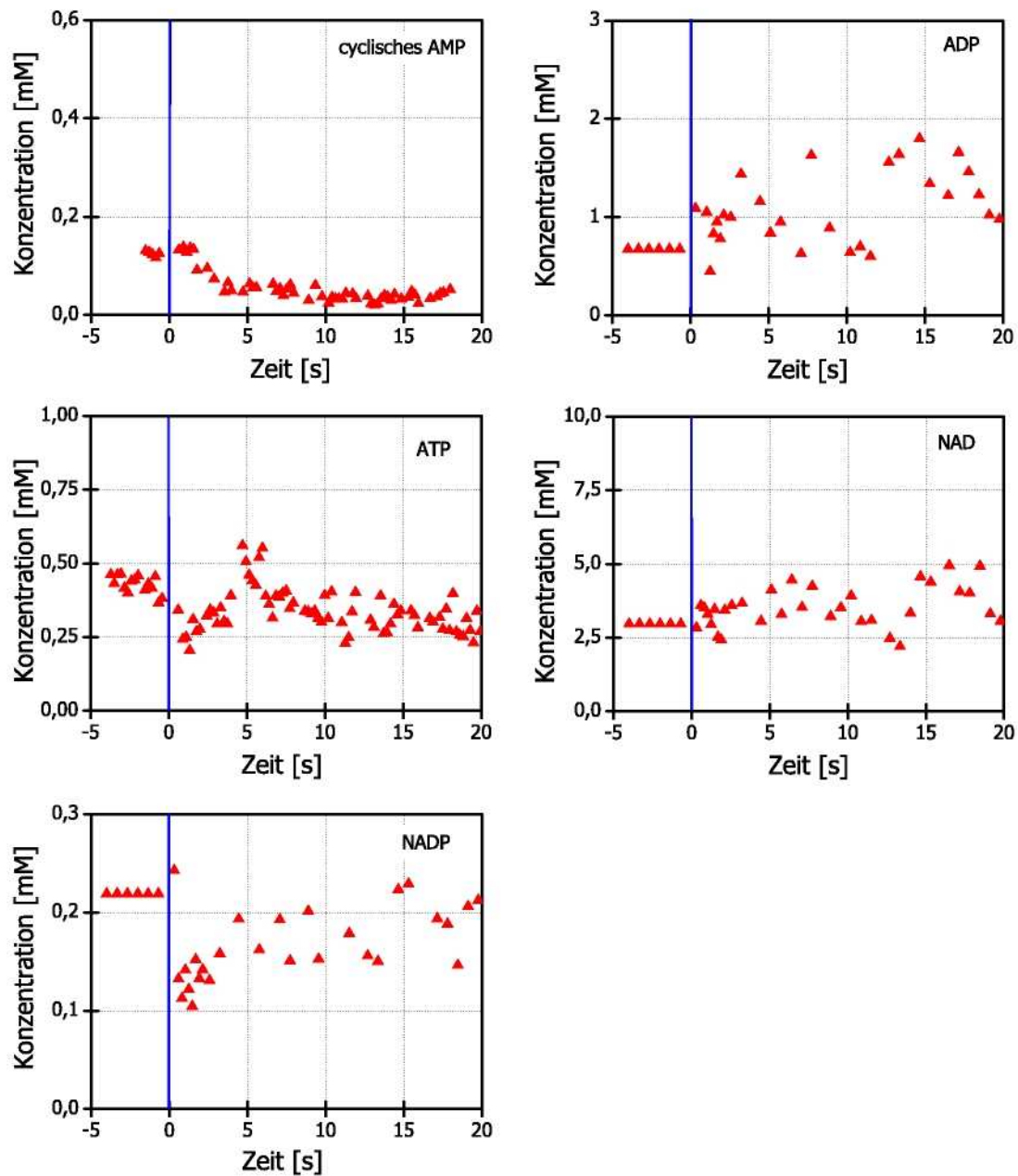


Abb. C-4 Trajektorien der Nukleotide aus Experiment WT1; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

C.2 Experiment MU1

Abbildung C–5 zeigt die Intermediate der verschiedenen Stoffwechselwege an, die im Experiment MU1 quantifiziert werden konnten, s. (Oldiges 2004). In den Abbildungen C–6, C–7 und C–8 ist der zeitliche Verlauf der Metabolitkonzentrationen vor und nach dem Puls dargestellt.

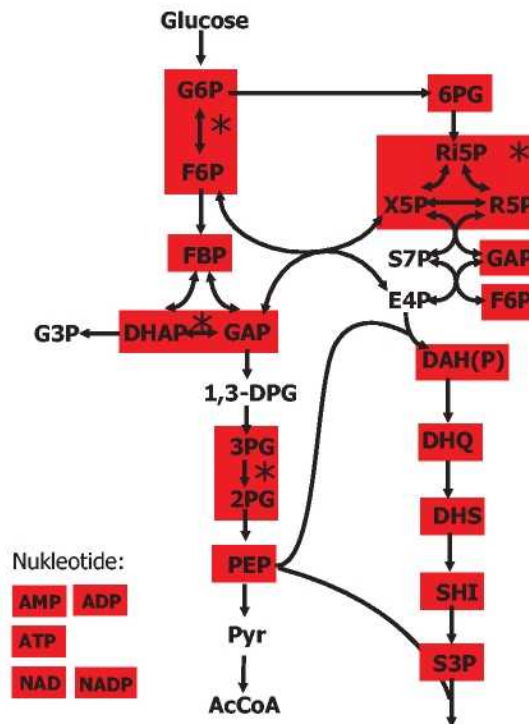


Abb. C–5 Überblick über die erfaßten Metabolite in Experiment MU1; eckige Kästen: mit der LC-MS-MS-Methode quantifizierte Metabolite; mit Sternchen gekennzeichnete Kästen: Anzeige von Metaboliten, deren Konzentration über die LC-MS-Methode nur in der Summe quantifiziert werden konnte

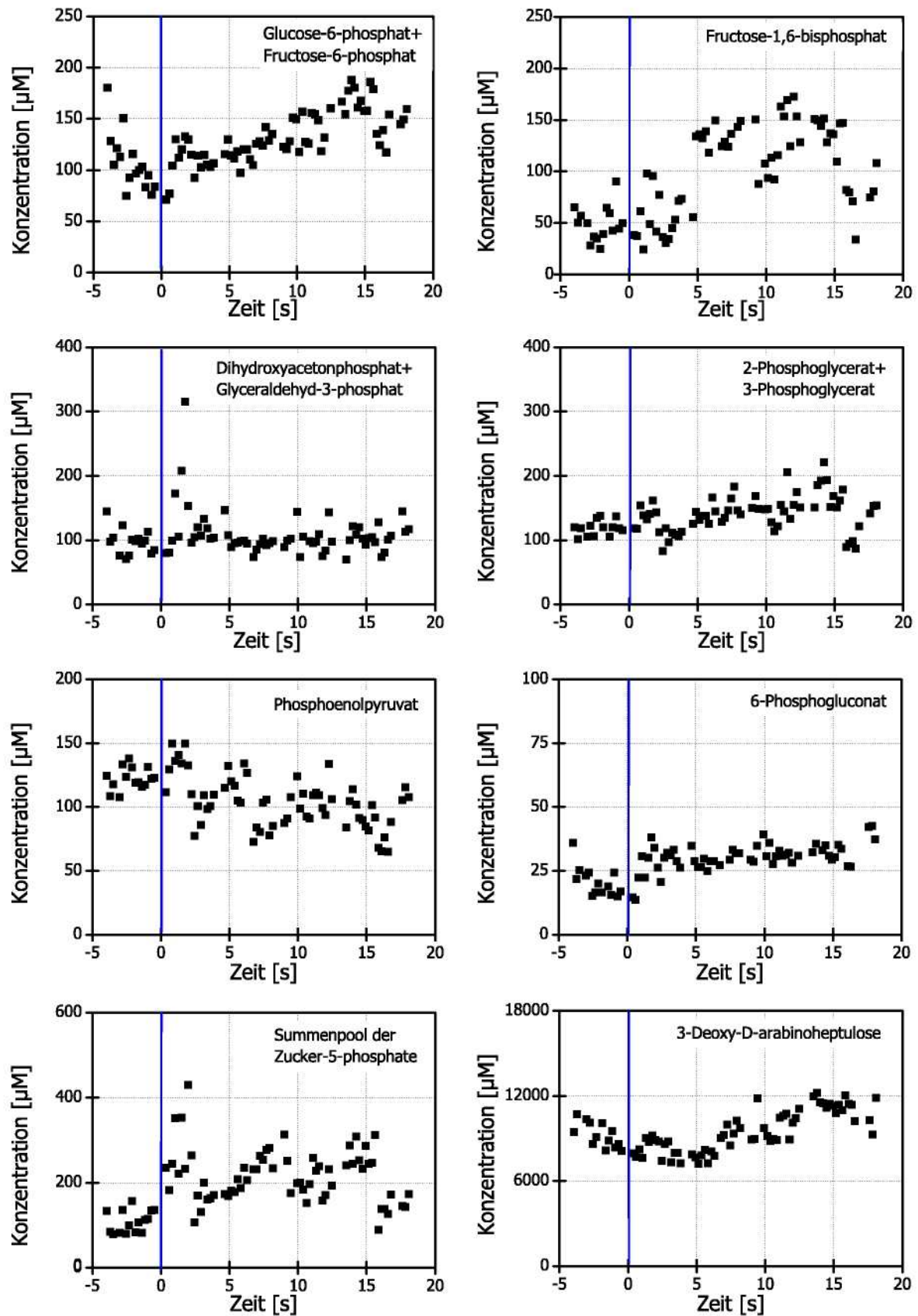


Abb. C-6 Meßdaten für Metabolite der Glykolyse, des Pentose-Phosphat-Weges und des Aromatenbiosyntheseweges aus Puls-Experiment MU1; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

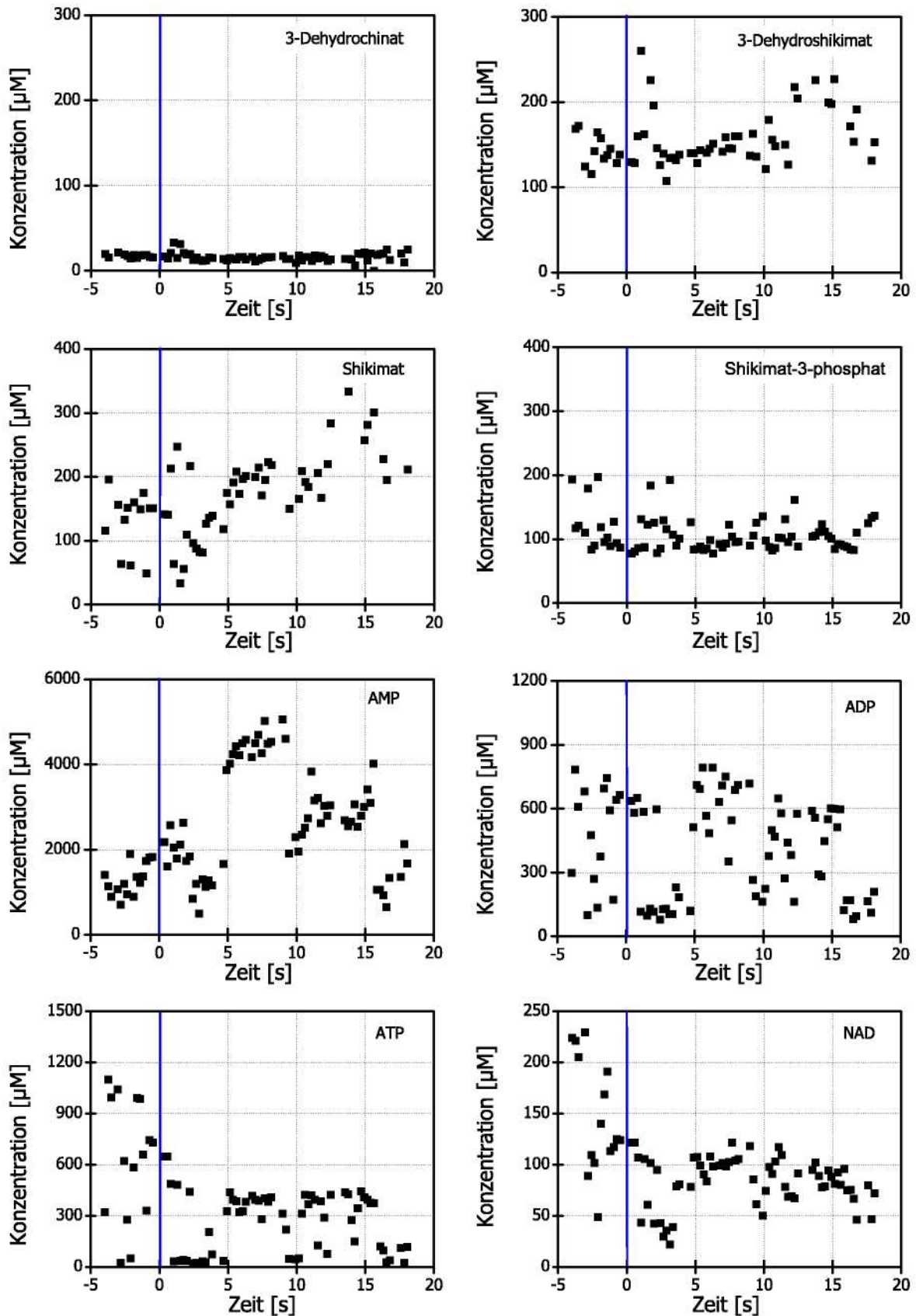


Abb. C-7 Meßdaten für Metabolite des Aromatenbiosyntheseweges und der Glykolyse, sowie die bestimmten Trajektorien einiger Nukleotide aus Puls-Experiment MU1; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

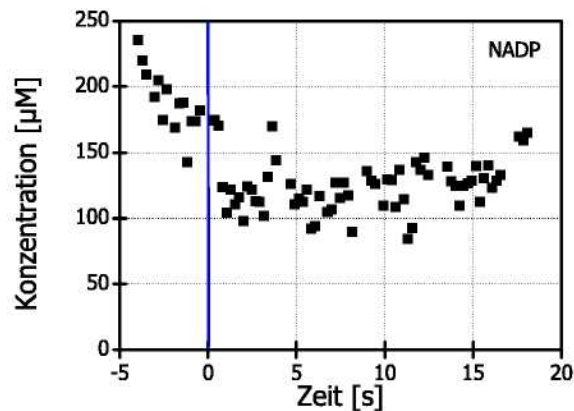


Abb. C–8 Verlauf von NAD in Pulsexperiment MU1; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

C.3 Experiment MU2

Abbildung C–9 zeigt die Intermediate der verschiedenen Stoffwechselwege an, die im Experiment MU2 quantifiziert werden konnten, s. (Oldiges 2004). In den Abbildungen C–10, C–11 und C–12 ist der zeitliche Verlauf der Metabolitkonzentrationen vor und nach dem Puls dargestellt.

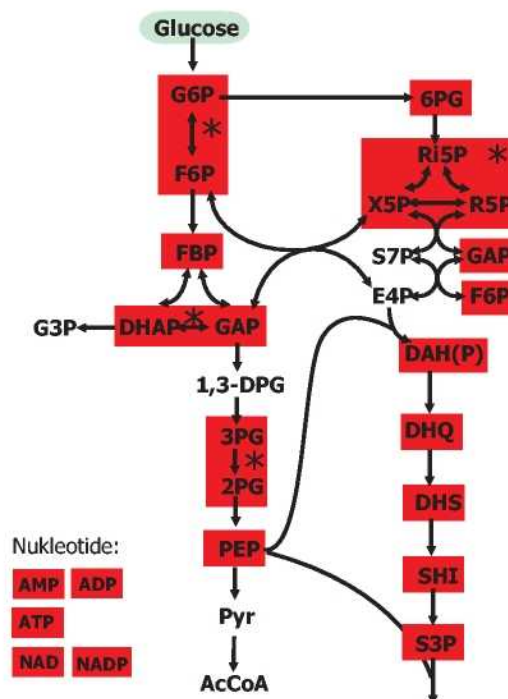


Abb. C–9 Überblick über die erfaßten Metabolite in Experiment MU2; rote Kästchen: mit der LC-MS-MS Technik, abgerundete Kästchen: mit enzymatischen Assays quantifizierte Metabolite; mit Sternchen gekennzeichnete Kästen zeigen die Metabolite an, deren Konzentration über die LC-MS-Methode nur in der Summe bestimmt werden konnte.

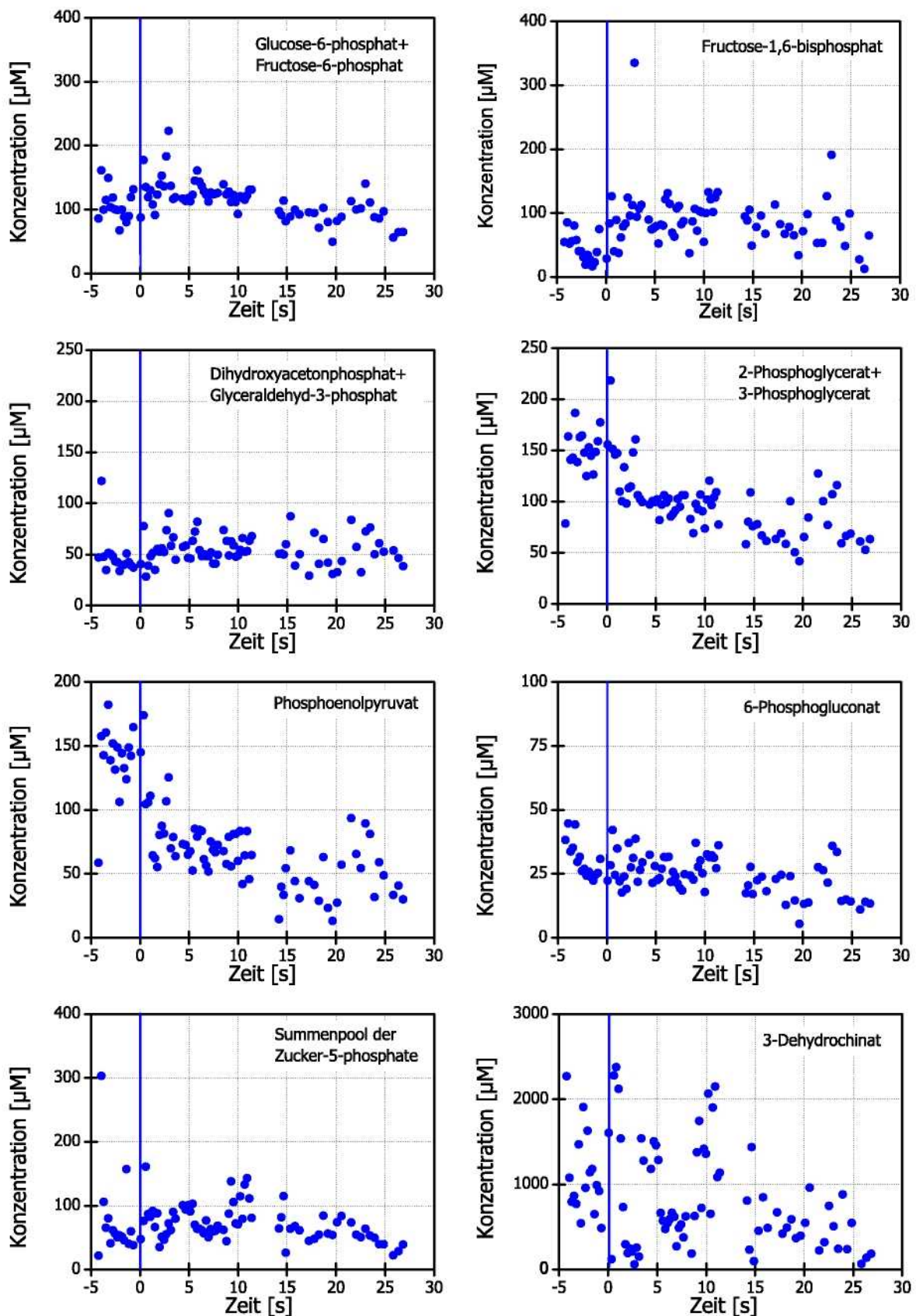


Abb. C–10 Meßdaten für Metabolite der Glykolyse, des Pentose-Phosphat-Weges und des Aromatenbiosyntheseweges aus Puls-Experiment MU2; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

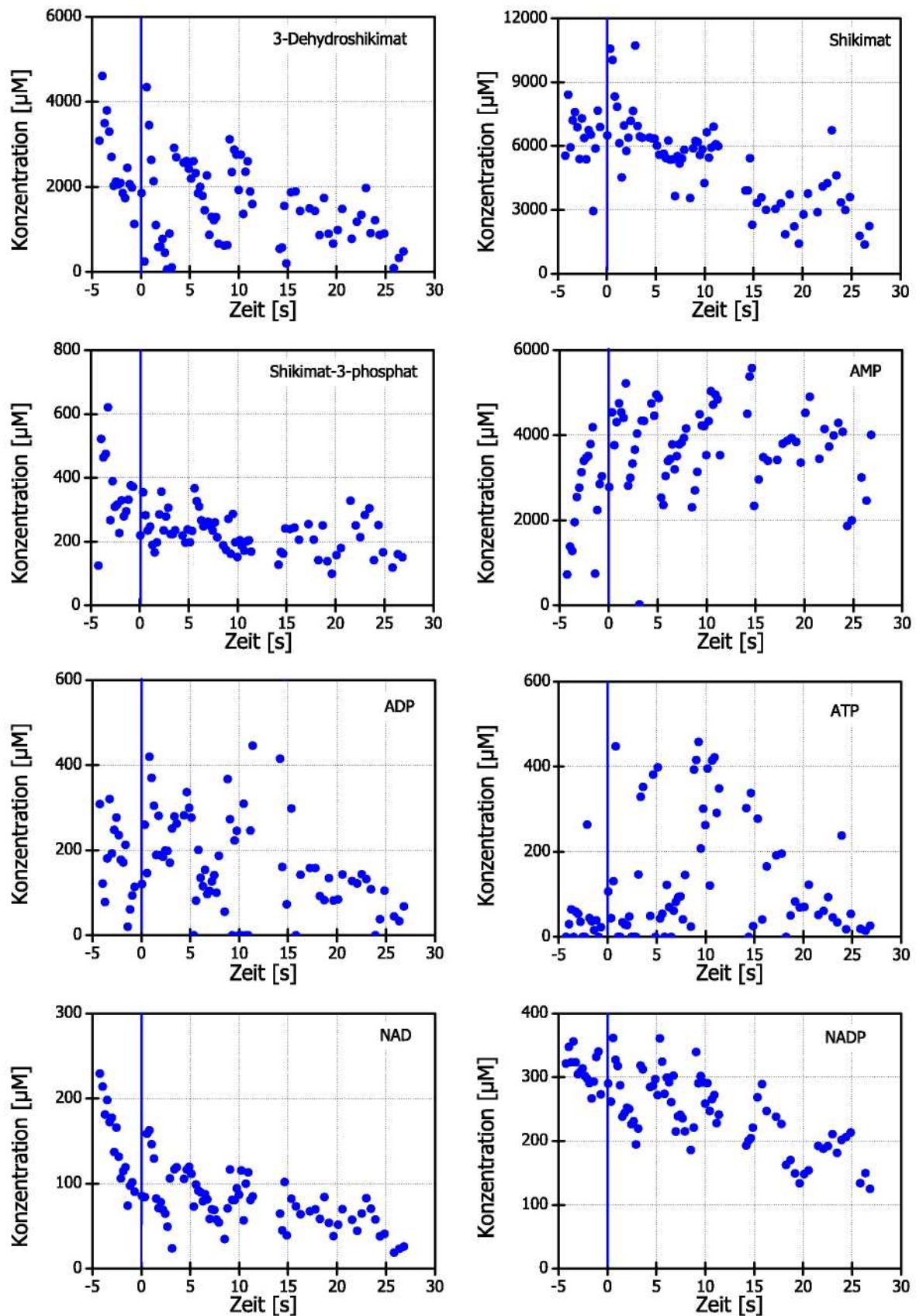


Abb. C-11 Meßdaten der Metabolite des Aromatenbiosyntheseweges und der Glykolyse, sowie die bestimmten Trajektorien einiger Nukleotide aus Puls-Experiment MU2; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

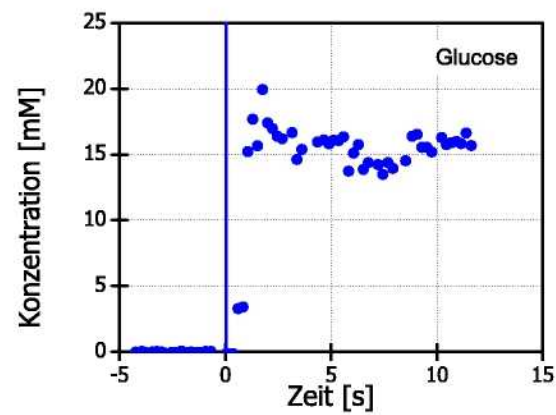


Abb. C-12 Verlauf von NAD und Glucose (extrazellulär) in Pulsexperiment MU2; die senkrechte Linie bei 0 s zeigt den Zeitpunkt des Glucose-Pulses an

Anhang D Enzymkinetiken

D.1 Allgemeines

Die Henri-Michaelis-Menten Gleichung, die als erste Gleichung zur Beschreibung von Enzymkinetiken (Henri 1902; Michaelis und Menten 1913) verwendet wurde, kann aus dem folgenden molekularen Reaktionsmechanismus abgeleitet werden:



Der Reaktionsmechanismus geht also von der reversiblen Bildung eines Enzym-Substrat-Komplexes (ES) aus, der in das Enzym E und Produkt P zerfällt. Unter der Annahme, daß die Bildung von (ES) schnell verläuft und damit ein der Zersetzungsreaktion vorgelagertes Gleichgewicht darstellt, ergibt sich die Gleichung:

$$v = \frac{dP}{dt} = -\frac{dS}{dt} = \frac{k_2 \cdot E_0 \cdot S}{K_m + S} \quad \text{mit} \quad K_m = \frac{k_{-1}}{k_1}, \quad (D-2)$$

wobei E_0 die Konzentration des freien Enzyms bezeichnet. Später wurde festgestellt, daß Bildung und des Zerfall des (ES)-Komplexes schnell verlaufen und daher eher die Annahme der Quasi-Stationarität bzgl. der (ES)-Konzentration getroffen werden sollte (Cornish-Bowden 2004). Dadurch bleibt die Form der Kinetik erhalten, während die Michaelis-Menten Konstante anders definiert ist: $K_m = (k_{-1} + k_2)/k_1$. k_2 wird häufig als k_{cat} oder Turnover-Zahl bezeichnet, d.h. die Konstante gibt die Zahl an Produktmolekülen an, die pro Zeiteinheit aus einem Substrat erzeugt werden. Die molare Enzymkonzentration E_0 ist bei Experimenten meist nicht exakt bekannt, weil das Enzym evtl. nicht vollständig aufgereinigt wurde, die molare Masse nicht bekannt ist oder Anteile des aufgereinigten Enzym katalytisch nicht mehr aktiv sind. Daher werden k_{cat} und E_0 in Gleichung (D-2) zur einer Grenzggeschwindigkeit, der maximalen Reaktionsgeschwindigkeit V_{max} , zusammengefaßt. Ebenso werden die bei den Experimenten eingesetzten Enzymmengen häufig nicht in molaren oder Massenkonzentrationen angegeben, sondern in Einheiten, die die katalytische Aktivität des Enzymes ausdrücken: eine „Unit“ (U) bezeichnet die Menge an eingesetzten Enzym, die ein μmol Substrat in einer Minute unter definierten Bedingungen umsetzen kann.

K_m gibt die Substratkonzentration wieder, bei der die Hälfte der maximalen Reaktionsgeschwindigkeit V_{max} , erreicht wird (Halbsättigungskonstante), und ist aus verschiedenen Gründen interessant: sie ist z.B. ein wichtiger Indikator für die Substrataffinität eines Enzyms und gibt einen näherungsweise gültigen Wert für die intrazelluläre Konzentration des Substrates an, da sehr viel kleinere und größere Konzentrationen für Vorgänge zur Regulation der Enzymaktivität wenig sinnvoll wären.

Lineweaver und Burk (Lineweaver und Burk 1934) konvertierten Gleichung (D-2) in eine lineare Form, um bei einer graphischen Auswertung genauere Werte für die Parameter bestimmen zu können:

$$\frac{1}{v} = \frac{1}{V_{max}} + \left(\frac{K_m}{V_{max}} \right) \frac{1}{S}. \quad (D-3)$$

Ein doppelt-reziproker Auftrag der Geschwindigkeit über der Substratkonzentration ergibt also eine Gerade, wobei der Wert von K_m bzw. V_{max} aus dem Schnittpunkt mit der Abszisse bzw. Ordinate entnommen werden kann. Daneben sind noch weitere Methoden zur Linearisierung der Enzymkinetiken bekannt, wie z.B. der Hanes- (S/v vs S) oder Eadie-Hofstee-Plot (v vs v/S).

D.2 Einfache Inhibitionskinetiken

Jedes Molekül, das die Geschwindigkeit der Enzymkatalyse reduziert, wird als Inhibitor betrachtet. Hier werden kurz die verschiedenen Formen der Inhibitionen anhand von Enzymen, die ein Substrat umsetzen und von einem Inhibitor beeinflusst werden, dargestellt.

Substratinhibition Bei hohen Substratkonzentrationen wird die Enzymaktivität häufig negativ durch den Überschuß an Substrat beeinflusst, indem der (ES)-Komplex reversibel ein zweites Substratmolekül bindet und dadurch inaktiv wird. Wird ansonsten der molekulare Reaktionsmechanismus zur Ableitung der Michaelis-Menten-Gleichung zugrundegelegt, s. Abbildung D–1, ergibt sich die folgende Kinetik:

$$v = \frac{k_2 \cdot E_0 \cdot S}{1 + K_m/S + S/K_i} = \frac{V_{\max} S}{S(1 + S/K_i) + K_m} \quad \text{mit} \quad K_i = \frac{k_{-i}}{k_i}. \quad (\text{D-4})$$

Die Geschwindigkeit der Enzymkatalyse durchläuft also ein Maximum, dessen Lage durch $S_{\max} = (K_i K_m)^{1/2}$ festgelegt wird.

Kompetitive Inhibition Ein kompetitiver Inhibitor ist ein Molekül mit einer dem Substrat ähnlichen Struktur oder Eigenschaften, wie z.B. ein Reaktionsprodukt oder ein nicht-metabolisierbares Substratderivat. Dieses konkurriert mit dem Substrat um die aktiven Zentren des Enzyms, in denen die Reaktion katalysiert wird, und bildet reversibel einen Enzym-Inhibitor-Komplex, der sich im Gleichgewicht mit dem freien Enzym befindet. Bei Erweiterung des Reaktionsmechanismus (D–1) um diese Reaktion läßt sich die folgende Gleichung ableiten:

$$v = \frac{V_{\max} \cdot S}{K_m(1 + I/K_i) + S}, \quad (\text{D-5})$$

wobei I für die Konzentration des freien Inhibitors steht. Die Gleichung kann in Form einer Michaelis-Menten-Gleichung wiedergegeben werden,

$$v = \frac{V_{\max}^{\text{app}} \cdot S}{K_m^{\text{app}} + S} \quad \text{mit} \quad V_{\max}^{\text{app}} = V_{\max} \quad \text{und} \quad K_m^{\text{app}} = K_m(1 + I/K_i), \quad (\text{D-6})$$

deren Wert für die maximale Reaktionsgeschwindigkeit $V_{\max}^{\text{app}}$ erhalten bleibt, während der Wert der Michaelis-Menten-Konstante K_m^{app} sich in Bezug auf die nicht-inhibierte Reaktion vergrößert. Aus der Formel ist weiterhin erkennbar, daß die kompetitive Inhibition durch hohe Substratkonzentrationen unterdrückt werden kann.

Nicht-kompetitive Inhibition Nicht-kompetitive Inhibitoren binden reversibel an allosterische Zentren außerhalb der aktiven Zentren des Enzyms und verursachen Konformationsänderungen, die die enzymatische Aktivität senken. Der Inhibitor kann somit an freie Formen des Enzymes und den (ES)-Komplex binden, s. Abbildung D–1, was bei der Annahme von sich rasch einstellenden Gleichgewichten zu folgendem Geschwindigkeitsgesetz führt:

$$v = \frac{V_{\max} \cdot S}{S(1 + I/K_i) + K_m(1 + I/K_i)}. \quad (\text{D-7})$$

Hierbei bleibt der Wert K_m^{app} erhalten, während sich der Wert $V_{\max}^{\text{app}}$ um den Faktor $1/(1 + I/K_i)$ verkleinert.

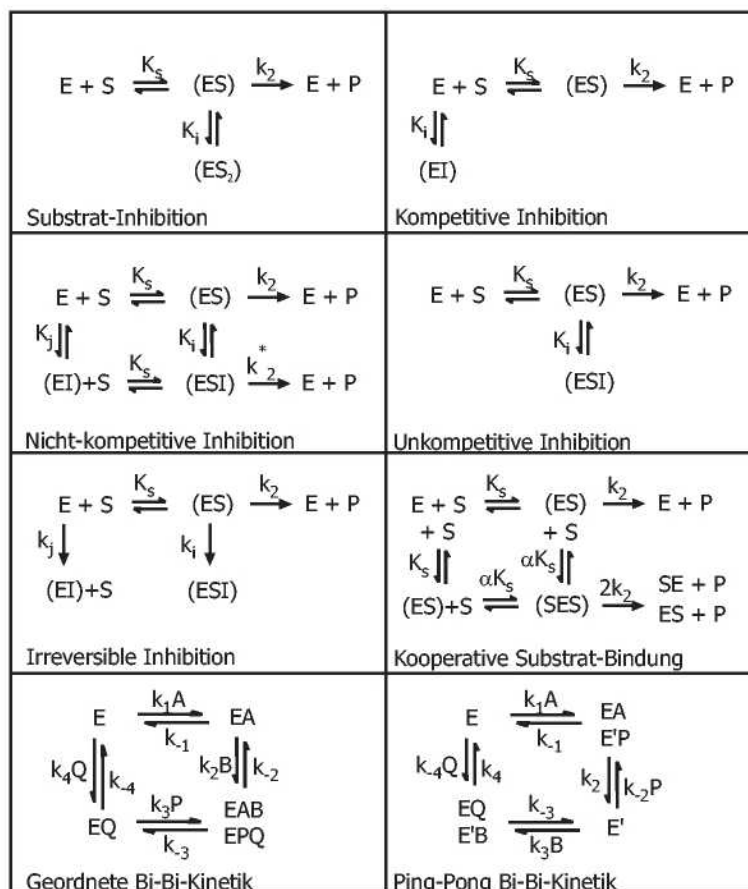


Abb. D–1 Molekulare Reaktionsmechanismen enzymatisch katalysierter Reaktionen

Unkompetitive Inhibition Unkompetitive Inhibitoren binden reversibel an einen (ES)-Komplex und inaktivieren ihn. Unter der Annahme, daß sich die (ESI)- und (ES)-Komplexe im Gleichgewicht befinden, ergibt sich die folgende Gleichung:

$$v = \frac{V_{\max} \cdot S}{K_m + S \cdot (1 + I/K_i)} \quad (D-8)$$

Die Substratinhibition ist also prinzipiell ein Sonderfall der unkompetitiven Inhibition. Bei dieser Form der Inhibition werden $V_{\max}$ - und K_m - der nicht-inhibierten Reaktion um den Faktor $1/(1 + I/K_i)$ gesenkt.

Irreversible Inhibition Die Inhibitoren verbinden sich kovalent mit dem Enzym zu einem Komplex, der nicht mehr dissoziiert, und können bei genügend großer Menge die Enzymaktivität auf Null absenken (Katalysatorgifte). Unter der Annahme, daß die kovalente Bindung schnell und vollständig erfolgt, wird durch eine bestimmte Menge irreversibler Inhibitoren eine äquivalente Menge an Enzymen inaktiviert. Die Katalyse durch die verbleibenden Enzyme gehorcht dann einer Michaelis-Menten-Kinetik mit einem reduzierten Wert für die maximale Reaktionsgeschwindigkeit: $V'_{\max} = k_2(E_0 - (ESI + EI))$.

D.3 Kooperativität allosterischer Enzyme

Allosterische Enzyme besitzen mehrere aktive Zentren, wobei die Bindung eines Moleküls strukturelle oder elektronische Veränderungen verursacht, die zu geänderten Substrataffinitäten der verbleibenden freien, aktiven Zentren führen. Solche Interaktionen können homotrop oder heterotrop sein, d.h. sie involvieren nur einen Substrattyp oder verschiedene Substrat- und Inhibitor-, bzw. Substrat- und Aktivatormoleküle. Meistens bestehen allosterische Enzyme aus mehreren Untereinheiten und ihre Geschwindigkeit-vs-Substrat-Kurven sind sigmoid. Der sigmoiden Kurvenform zufolge steigt die Enzymaktivität bei Phänomenen von positiver Kooperativität schneller als normal^{4.1}, bei Phänomenen von negativer Kooperativität langsamer als normal an. Kooperative Enzyme können auf Fluktuationen von Substrat- oder Effektorkonzentrationen unterschiedlich sensitiv reagieren, was physiologisch bedeutsam ist (z.B. An- oder Abstellen von Flüssen über das Enzym, keine Enzymkatalyse unterhalb eines Schwellenwertes).

Abbildung D–1 zeigt einen Reaktionsmechanismus, der einem positiven, homotropen Verhalten eines Enzymes mit zwei Bindungsstellen für das Substrat zugrundeliegt. Die Bindung des ersten Substratmoleküls erhöht die Substrataffinität des Enzyms für das zweite Molekül, was durch die Veränderung der Dissoziationskonstanten K_s des (ES)-Komplexes angezeigt wird: αK_s : $\alpha < 1$ bei positiver, $\alpha > 1$ bei negativer Kooperativität. Damit läßt sich das folgende Geschwindigkeitsgesetz ableiten, das nicht mehr mit der Michaelis-Menten-Gleichung in Übereinstimmung gebracht werden kann:

$$v = \frac{V_{\max} \left(\frac{S}{K_s} + \frac{S^2}{\alpha K_s^2} \right)}{1 + \frac{2S}{K_s} + \frac{S^2}{\alpha K_s^2}} \quad \text{mit} \quad K_s = \frac{E_0 \cdot S}{ES} \quad \text{und} \quad V_{\max} = 2k_2 E_0. \quad (\text{D-9})$$

Wenn das Enzym ein signifikant kooperatives Verhalten aufweist, ist der Faktor α sehr klein. Dies hat zur Folge, daß die Geschwindigkeit des Enzyms vorwiegend von den S^2 -Termen beeinflusst wird. Damit können die Konzentrationen aller Enzym-Substrat-Komplexe, die weniger als 2 Moleküle des Substrates beinhalten, bei Substratkonzentrationen, die höher als K_s sind, vernachlässigt werden. Gleichung (D–9) vereinfacht sich zu:

$$v = \frac{V_{\max} \frac{S^2}{K'}}{1 + \frac{S^2}{K'}}, \quad (\text{D-10})$$

bzw. allgemein für n äquivalente Substratbindungsstellen zu der Hill-Gleichung (Hill 1910):

$$v = \frac{V_{\max} S^n}{K' + S^n} \quad \text{mit} \quad K' = \alpha^{n-1} \cdot \beta^{n-2} \cdot \gamma^{n-3} \cdot \dots \cdot K_s^n. \quad (\text{D-11})$$

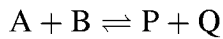
Statt K' wird auch häufig eine Hill-Konstante $K_{0,5}^n$ in die Hill-Gleichung eingeführt, die die Substratkonzentration angibt, bei der die Hälfte der maximalen Reaktionsgeschwindigkeit erreicht wird, und die eine Dimension von $[\text{mol l}^{-1}]$ (anstatt von $[(\text{mol l}^{-1})^n]$) besitzt.

D.4 Kinetiken mit mehr als einem Substrat

Fast alle enzymatischen Reaktionen, an denen zwei Substrate und zwei Produkte beteiligt sind, gehören zu den Gruppentransferreaktionen. Z.B. kann eine Phosphatgruppe durch die Hexokinase von ATP (A)

^{4.1}Normal bedeutet hier: bei Annahme einer Michaelis-Menten-Kinetik für die Geschwindigkeit der Enzymkatalyse.

auf Glucose (B) übertragen werden, wobei ADP (P) und G6P (Q) entstehen:



Solche Reaktionen können über einen ternären Komplex verlaufen, indem ein Enzym zunächst beide Substrate nacheinander bindet, sie zu den Produkten umsetzt und diese wiederum nacheinander in die Lösung abgibt. Gemäß der „induced-Fit“-Theorie von Koshland^{4,2} (Koshland Jr. et al. 1966) wird bei vielen Reaktionen eine bestimmte Reihenfolge für die Bindung der beiden Substrate und die Abgabe der Produkte eingehalten, s. Abbildung D–1. Aus diesem Reaktionsmechanismus kann die geordnete Bi-Bi-Kinetik abgeleitet werden:

$$v = \frac{V_f \left(AB - \frac{PQ}{K_{eq}} \right)}{AB \left(1 + \frac{P}{K_{ip}} \right) + K_{mB}(A + K_{iA}) + K_{mA}B + T} \quad (D-12)$$

$$T = \frac{V_f}{V_r K_{eq}} \left[K_{mQ}P \left(1 + \frac{A}{K_{iA}} \right) + Q \left(K_{mp} \left(1 + \frac{K_{mA}B}{K_{iA}K_{mB}} \right) + P \left(1 + \frac{B}{K_{iB}} \right) \right) \right],$$

mit der maximalen Reaktionsgeschwindigkeit der Hin- bzw. Rückreaktion V_f bzw. V_r , der Gleichgewichtskonstanten K_{eq} , den Michaelis-Menten Konstanten für A, B, P und Q, K_{mA} , K_{mB} , K_{mP} , K_{mQ} , (K_{mA} gibt die Konzentration von A wieder, für die $v = V_f/2$ ist, wenn das Enzym mit B gesättigt ist, wobei P und Q nicht vorliegen, usw.) und den Produktinhibitionskonstanten, K_{iA} , K_{iB} , K_{iP} , K_{iQ} (K_{iA} gibt die Inhibierungskonstante von A auf die Rückreaktion wieder, usw.).

Ein alternativer Reaktionsmechanismus geht nicht von einem ternären Komplex aus, sondern nimmt an, daß zunächst der Gruppendonator gebunden wird, der die funktionelle Gruppe an das Enzym überträgt, und im Anschluß in die Lösung freigelassen wird. Danach erfolgt die Bindung des Gruppenakzeptor, der die funktionelle Gruppe vom Enzym E' übernimmt, s. Abbildung D–1. Aus einem solchen Reaktionsmechanismus kann die Ping-Pong-Bi-Bi-Kinetik abgeleitet werden:

$$v = \frac{V_f \left(AB - \frac{PQ}{K_{eq}} \right)}{AB + K_{mB}A + K_{mA}B \left(1 + \frac{Q}{K_{iQ}} \right) + T} \quad (D-13)$$

$$T = \frac{V_f}{V_r K_{eq}} \left[K_{mQ}P \left(1 + \frac{A}{K_{iA}} \right) + Q(K_{mp} + P) \right].$$

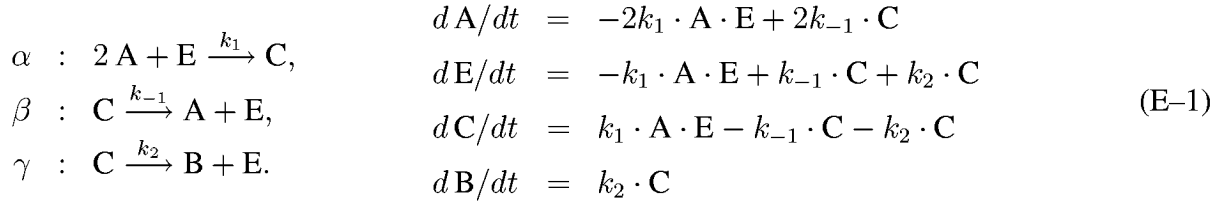
Die Bedeutung der Konstanten ist unter Gleichung (D–12) erläutert.

Auf ähnliche Weise sind auch komplexere Reaktionsmechanismen für Reaktionen mit mehr als zwei Substraten oder mehreren Substraten und Effektoren aufgestellt worden. Die King-Altman-Methode (King und Altman 1956) ermöglicht dabei eine relativ einfache, schematische Ableitung der Geschwindigkeitsgesetze aus den zugehörigen Reaktionsmechanismen.

^{4,2}Im Gegensatz zu Fischer's „Schlüssel-Schloß-Theorie“ (Fischer 1894), die annimmt, daß das aktive Zentrum eines Enzyms immer die zur Substrat-Bindung benötigte Paßform besitzt, geht diese Theorie von der Flexibilität des aktiven Zentrums aus. D.h. das aktive Zentrum hat zwar generell das Potential, ein bestimmtes Substrat genau einzupassen, nimmt die dafür notwendige Konformation jedoch erst unmittelbar vor der Substratbindung (z.B. durch die Bindung anderer Substrate) ein.

Anhang E Vereinfachungen des Reaktionsnetzwerkes

Aufgrund der Berücksichtigung von Massenbilanzen sind Vereinfachungen der Reaktionsnetzwerke möglich, die am Beispiel des folgenden chemischen Reaktionssystems dargestellt werden:



Zur Vereinfachung Beispiel-Reaktionsnetzwerkes bietet es sich an, die Reaktionen β und γ bzgl. des Reaktanden C zur Reaktion β' zusammenzufassen:



Die Edukte und Produkte der Reaktion β' ergeben sich demzufolge als Schnittmenge der Reaktanden der vorherigen Reaktionen β und γ . Der Term für die Reaktionsgeschwindigkeit $r_{\beta'}$ wird über Zusammenfassen entsprechender Terme in Gleichung (E-1) erhalten:

$$r_{\beta'} = r_{\beta} + r_{\gamma} = (k_{-1} + k_2)C,$$

bzw. allgemein:

$$r_{\alpha'j} = \sum_{i=1}^k r_{\alpha i}, \quad (E-2)$$

wobei r die Ausdrücke für die Kinetiken der Reaktionen α und k die Anzahl der zusammengefaßten Reaktionen bezeichnet.

Zur Berechnung der geänderten stöchiometrischen Koeffizienten der Reaktanden in Reaktion $r_{\alpha'}$ wird die allgemeine Beschreibung einer Differentialgleichung für die zeitabhängige Konzentrationsänderung eines Reaktanden herangezogen^{5.1}:

$$\frac{dC_j}{dt} = \sum_{\alpha=1}^m r_{\alpha} n_{\alpha j}, \quad j = 1, \dots, n. \quad (E-3)$$

m gibt die Zahl der Reaktionen, n die Zahl der Reaktanden, und $n_{\alpha j}$ den stöchiometrischen Koeffizienten der Reaktion r_{α} , an der Reaktand C_j beteiligt ist, wieder. Damit läßt sich für eine bzgl. des Reaktanden C_j zusammengefaßte Reaktion α' der stöchiometrische Koeffizient $n_{\alpha'j}$ ableiten:

$$r_{\alpha'} \cdot n_{\alpha'j} = \sum_{i=1}^k n_{ij} r_{\alpha i} = r_{\alpha'} \cdot \sum_{i=1}^k n_{ij} \left(\frac{r_{\alpha i}}{r_{\alpha'}} \right) \Rightarrow n_{\alpha'j} = \sum_{i=1}^k n_{ij} \left(\frac{r_{\alpha i}}{r_{\alpha'}} \right). \quad (E-4)$$

^{5.1}Diese ist in Gleichung (E-1) für das Beispielsystem konkretisiert.

Unter Verwendung von Gleichung (E-4) ergeben sich z.B. die stöchiometrischen Koeffizienten $n_{\beta'C}$ bzw. $n_{\beta'A}$ des Beispielsystems zu:

$$n_{\beta'C} = \frac{n_{\beta C} \cdot r_{\beta} + n_{\gamma C} \cdot r_{\gamma}}{r_{\beta'}} = \frac{(1 \cdot k_{-1} + 1 \cdot k_2)C}{(k_{-1} + k_2)C}, \quad \text{d.h.} \quad n_{\beta'C} = 1,$$

$$n_{\beta'A} = \frac{n_{\beta A} \cdot r_{\beta} + n_{\gamma A} \cdot r_{\gamma}}{r_{\beta'}} = \frac{(2 \cdot k_{-1} + 0 \cdot k_2)C}{(k_{-1} + k_2)C}, \quad \text{d.h.} \quad n_{\beta'A} = \frac{2k_{-1}}{k_{-1} + k_2}.$$

Eine notwendige Voraussetzung für die beschriebene Transformation ist, daß die stöchiometrischen Koeffizienten $n_{\alpha'j}$ über die Zeit hinweg konstant bleiben, also keine Funktionen der Konzentrationen der Reaktanden darstellen, $r_{\alpha j}/r_{\alpha'} = p_i = \text{konst.} \forall i = 1, \dots, k$.

Auf eine ähnliche Weise können Reaktanden in verschiedenen Reaktionen bzw. Reaktanden und Reaktionen zusammengefaßt werden, s. (Zeigler 1988) für eine Abschätzung der bei diesen Kombinationen begangenen Fehlern. Tabelle E-1 zeigt die für die Transformationen notwendigen Operationen an, wobei der Quotient R_{α}/π bezeichnet, daß π Reaktanden R_{α} zu einem Reaktanden R'_{α} zusammengezogen wurden.

Tab. E-1 Vereinfachungen des Reaktionsnetzwerkes

Spezifikation des Reaktionsnetzwerkes								
Reaktion	Reaktanden	Produkte	Geschwindigkeiten	Stöchiometrische Koeffizienten				
α_1	R_{α_1}	P_{α_1}	r_{α_1}	n_{11}	n_{12}	n_{13}	$\dots$	n_{1n}
α_2	R_{α_2}	P_{α_2}	r_{α_2}	n_{21}	n_{22}	n_{23}	$\dots$	n_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$
α_m	R_{α_m}	P_{α_m}	r_{α_m}	n_{m1}	n_{m2}	n_{m3}	$\dots$	n_{mn}
Vereinfachtes Reaktionsnetzwerk durch Verschmelzen von Reaktionen								
Reaktion	Reaktanden	Produkte	Geschwindigkeiten	Stöchiometrische Koeffizienten				
$\alpha' = \sum_{i=1}^k \alpha_i$	$R_{\alpha'} = \bigcup_{i=1}^k R_{\alpha_i}$	$P_{\alpha'} = \bigcup_{i=1}^k P_{\alpha_i}$	$r_{\alpha'} = \sum_{i=1}^k r_{\alpha_i}$	$n_{\alpha'j} = \sum_{i=1}^k n_{ij} \left(\frac{r_{\alpha i}}{r_{\alpha'}} \right) \quad \dots$				
α_{k+1}								
$\vdots$								
α_n								
Vereinfachtes Reaktionsnetzwerk durch Verschmelzen von Reaktanden								
Reaktion	Reaktanden	Produkte	Geschwindigkeiten	Stöchiometrische Koeffizienten				
α_1	$R'_{\alpha} = R_{\alpha 1}/\pi$	$P'_{\alpha} = P_{\alpha 1}/\pi$	r_{α}	$n'_{\alpha,1} = \sum_{j=1}^l n_{\alpha,j}$				
	$R_{\alpha_{l+1}}$	$P_{\alpha_{l+1}}$						
	$\vdots$	$\vdots$						
	R_{α_n}	P_{α_n}						
Vereinfachtes Reaktionsnetzwerk durch Verschmelzen von Reaktanden und Reaktionen								
Reaktion	Reaktanden	Produkte	Geschwindigkeiten	Stöchiometrische Koeffizienten				
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$			
α'	$R_{\alpha'} = \bigcup_{\alpha \in (\alpha')} R_{\alpha}/\pi$	$P_{\alpha'} = \bigcup_{\alpha \in (\alpha')} P_{\alpha}/\pi$	$r'_{\alpha} = \sum_{\alpha \in (\alpha')} r_{\alpha}$	$n_{\alpha'j'} = \sum_{\alpha \in (\alpha')} (r_{\alpha}/r_{\alpha'}) \times \sum_{j \in (j')} n_{\alpha,j}$				

Anhang F Simulation und Optimierung

F.1 Numerische Integration

Die Darstellung der numerischen Integration folgt im Wesentlichen (Deuffhard und Bornemann 2002; Hanke Bourgeois 2002).

F.1.1 Allgemeines

Ein Anfangswertproblem (AWP) wird mathematisch folgendermaßen beschrieben:

$$\dot{\mathbf{y}} = \frac{d\mathbf{y}(t)}{dt} = f(t, \mathbf{y}(t)), \quad \mathbf{y}(t_0) = \mathbf{y}_0. \quad (\text{F-1})$$

Aufgabe ist es, zu einem vorgegebenen $t_b \neq t_0$ durch Integration von $f(t, \mathbf{y}(t))$ eine Lösung $\mathbf{y}(t_b)$ zu berechnen, wobei die Werte von $\mathbf{y}(t)$ zum Zeitpunkt t_0 als $\mathbf{y}_0$ gegeben sind, und $t_b > t_0$ sein soll.

Dazu ist zunächst von Interesse, ob eine Lösung des AWP existiert und eindeutig bestimmt ist. Dies garantiert in den meisten Fällen der Eindeutigkeits- und Existenzsatz.

Satz F.1.1 (Picard-Lindelöf) *Ist die Funktion f in t stetig und erfüllt eine Lipschitz-Bedingung, d.h. es existiert eine Konstante $L \geq 0$, so daß gilt:*

$$\|f(t, \mathbf{y}) - f(t, \mathbf{x})\| \leq L \|\mathbf{y} - \mathbf{x}\|, \quad \forall t \in [t_0, t_b], \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

dann gibt es eine eindeutige Lösung $\tilde{\mathbf{y}}(t)$ des in Gl. (F-1) definierten AWP.

Häufig kann eine allgemeine Lösung des AWP explizit nicht berechnet werden. In diesen Fällen ist man auf numerische Näherungsverfahren angewiesen, die im Folgenden besprochen werden.

Die numerischen Integrationsverfahren arbeiten mit einer *Diskretisierung*, d.h. anstelle einer kontinuierlichen Lösung $\mathbf{y}(t), t_0 \leq t \leq t_b$ wird eine Zerlegung des Integrationsintervalls: $t_0 = t_1 < t_2 \dots < t_m = t_b$, mit den Näherungen $\mathbf{Y}_j, j = 1, \dots, m$ der Lösung $\mathbf{y}(t_j)$ betrachtet. Die t_j werden *Integrationsknoten*, $I_h = \{t_0, \dots, t_m\}$ *Integrationsgitter*, die $h_j := t_{j+1} - t_j, j = 0, \dots, m-1$ *Schrittweiten* und $\delta := \max h_j$ die *Feinheit* des Integrationsgitters I_h genannt. Damit lassen sich die Näherungen auch als Funktionswerte einer sogenannten *Gitterfunktion*: $Y_h : I_h \rightarrow \mathbb{R}^n$, interpretieren. Ein *Diskretisierungsverfahren* ist dann ein Verfahren, das jedem Gitter I_h eine Gitterfunktion Y_h zuordnet. Die numerischen Verfahren zur Integration gewöhnlicher Differentialgleichungen werden üblicherweise in die Klassen der Einschritt-, Mehrschritt- und Extrapolationsverfahren unterteilt.

Einschrittverfahren verwenden jeweils nur die zuletzt berechnete Näherung $(t_j, \mathbf{Y}_j)$ um die folgende Näherung $(t_{j+1}, \mathbf{Y}_{j+1})$ zu bestimmen. Sie haben die Form:

$$\mathbf{Y}_{j+1} = \mathbf{Y}_j + h_j \Phi(t_j, \mathbf{Y}_j; h_j). \quad (\text{F-2})$$

Die Funktion Φ heißt *Verfahrens-* oder *Inkrementfunktion* des jeweiligen Verfahrens und gibt die von der rechten Seite der Gleichung (F-1) abhängige „Fortschreiterichtung“ der Integration wieder. Das Produkt $h_j \Phi$ wird auch als *Quadraturformel* bezeichnet.

Mehrschrittverfahren verwenden mehrere bereits berechnete Näherungen $(t_i, \mathbf{Y}_i), j < i < j+s$, um die Lösung $(t_{j+s}, \mathbf{Y}_{j+s})$ zu bestimmen. Bei Verwendung von linearen Ansätzen für die f_i - bzw. $\mathbf{Y}_i$ -Daten

und äquidistanten Schrittweiten wird die folgende Gleichung für ein lineares Mehrschrittverfahren erhalten:

$$\sum_{i=0}^s \alpha_i \mathbf{Y}_{j+i} = h \sum_{i=0}^s \beta_i f_{j+i}. \quad (\text{F-3})$$

Dabei sind $\alpha_i, \beta_i \in \mathbb{R}, i = 0, \dots, s, \alpha_s \neq 0, t_i := a + ih, h := (b-a)/m$ und $f_i := f(t, \mathbf{Y}_i), i = 0, 1, \dots$. Ist $\beta_s = 0$, so läßt sich Gleichung (F-3) nach $\mathbf{Y}_{j+s}$ auflösen, d.h. das Verfahren ist *explizit*. Ist $\beta_s \neq 0$, so stellt F-3 eine implizite Gleichung zur Berechnung von $\mathbf{Y}_{j+s}$ dar, und das Verfahren wird *implizit* genannt.

Extrapolationsverfahren beruhen auf einem Einschritt- bzw. Mehrschrittverfahren, wobei Näherungen für $\mathbf{y}(t_b)$ zu verschiedenen Schrittweiten berechnet werden. Diese Näherungen werden durch Extrapolation bzgl. der Schrittweite verbessert.

F.1.2 Einschrittverfahren

Vorstellung einiger Verfahren Das einfachste Verfahren ist das Polygonzug- oder Euler-Cauchy-Verfahren, das aus Gleichung (F-1) hervorgeht, indem die Ableitung f durch den Vorwärtsdifferenzenquotienten $(\mathbf{Y}_{j+1} - \mathbf{Y}_j)/h_j$ ersetzt wird:

$$\mathbf{Y}_{j+1} = \mathbf{Y}_j + h_j f(t_j, \mathbf{Y}_j). \quad (\text{F-4})$$

Beinhaltet das DGL-System nur eine abhängige Variable y , läßt sich die in Gleichung (F-4) angegebene Näherung des Integralwertes anschaulich als Rechteck mit der Schrittweite h als Länge und dem Funktionswert am linken Rand $f(t_j, \mathbf{Y}_j)$ als Höhe interpretieren. Wird anstelle des Vorwärts- der Rückwärtsdifferenzenquotient verwendet, resultiert das *implizite* Euler-Cauchy-Verfahren: $\mathbf{Y}_{j+1} = \mathbf{Y}_j + h_j f(t_{j+1}, \mathbf{Y}_{j+1})$.

Über die geometrische Interpretation der Quadraturformel können Verbesserungen des Verfahrens abgeleitet werden. Z.B. kann der Integralwert im eindimensionalen Fall auch durch ein Trapez anstelle eines Rechtecks angenähert werden. Das *Verfahren nach Heun* verwendet z.B. eine abgewandelte Sehnentrapez- als Quadraturformel bzw. mittelt zwei Richtungsfelder zur Erzeugung der Inkrementfunktion:

$$\begin{aligned} k_1 &:= f(t_j, \mathbf{Y}_j) \\ k_2 &:= f(t_{j+h}, \mathbf{Y}_j + h_j k_1) \\ \mathbf{Y}_{j+1} &:= \mathbf{Y}_j + h_j \left(\frac{k_1 + k_2}{2} \right) \end{aligned} \quad (\text{F-5})$$

Alternativ dazu könnte als Fortschreiterichtung auch eine mittlere Steigung gewählt werden, wie z.B. beim *modifizierten Euler-Verfahren* (Runge 1895):

$$\begin{aligned} k_1 &:= f(t_j, \mathbf{Y}_j) \\ k_2 &:= f(t_{j+\frac{1}{2}h}, \mathbf{Y}_j + h_j(\frac{1}{2}k_1)) \\ \mathbf{Y}_{j+1} &:= \mathbf{Y}_j + h_j k_2 \end{aligned} \quad (\text{F-6})$$

Direkte Verallgemeinerungen des Heunschen bzw. modifizierten Euler-Verfahrens sind die häufig verwendeten Runge-Kutta-Verfahren (RK-Verfahren), die die Verfahrensfunktion als Linearkombination von Auswertungen der Funktion $f(t, \mathbf{y}(t))$ in diskreten Punkten wählen. Die allgemeine Form eines s -stufigen Runge-Kutta-Verfahrens, die auf Arbeiten von Runge, Kutta und Heun um die Jahrhundertwende zurückgeht, s. z.B. (Kutta 1901), lautet:

$$\begin{aligned} \mathbf{Y}_{j+1} &= \mathbf{Y}_j + h_j \sum_{i=1}^s b_i k_i \\ k_i &= f \left(t_j + c_i h_j, \mathbf{Y}_j + h_j \sum_{l=1}^s a_{il} k_l \right). \end{aligned} \quad (\text{F-7})$$

Dabei bezeichnet $\mathbf{c} = (c_1, \dots, c_s)^T$ den Knotenvektor, $\mathbf{b} = (b_1, \dots, b_s)^T$ den Wichtungsvektor und $\mathbf{A} = (a_{il})$ die Verfahrensmatrix des jeweiligen Runge-Kutta-Verfahrens. Alle Koeffizienten, die das konkrete Verfahren festlegen, werden in einem Tableau, dem sogenannten *Butcher-Schema*, angeordnet, s. Tabelle F-1 (Butcher 1963). Wenn die Koeffizientenmatrix $\mathbf{A}$ eine strikt untere Dreiecksgestalt besitzt, also $a_{il} = 0 \forall l \geq i$, können die Hilfsgrößen k_i der Reihe nach explizit aus den Gleichungen berechnet werden (explizites RK-Verfahren), wie z.B. beim Heun- und modifizierten Euler-Verfahren, s. Tabelle F-1. Die Schemata von zwei weiteren, gebräuchlichen expliziten Runge-Kutta-Verfahren

Tab. F-1 Allgemeines Butcher-Schema (links); Butcher-Schema für das Heun- (rechts oben) bzw. modifizierte Euler-Verfahren (rechts unten).

$\begin{array}{c c} \mathbf{c} & \mathbf{A} \\ \hline & \mathbf{b} \end{array}$	c_1	a_{11}	a_{12}	$\dots$	a_{1s}	$\begin{array}{c cc} 0 & & \\ 1 & 1 & \\ \hline & 1/2 & 1/2 \end{array}$
	c_2	a_{21}	a_{22}	$\dots$	a_{2s}	
	$\vdots$	$\vdots$	$\vdots$		$\dots$	
	c_s	a_{s1}	a_{s2}	$\vdots$	$a_{s,s}$	
		b_1	b_2	$\dots$	b_s	
						$\begin{array}{c cc} 0 & & \\ 1/2 & 1/2 & \\ \hline & 0 & 1 \end{array}$

sind beispielhaft in Tabelle F-2 angegeben. Meistens sind RK-Verfahren jedoch implizit, so daß zur Berechnung i.A. k_i nichtlineare Gleichungssysteme zu lösen sind. Dabei werden zur Konstruktion eines RK-Verfahrens die Parameter des Butcher-Schemas so bestimmt, daß das Verfahren eine möglichst hohe Konsistenzordnung und gute Stabilitätseigenschaften, s. folgenden Abschnitt, besitzt.

Tab. F-2 Butcher-Schemata des klassischen RK-Verfahrens, RK 4, (links) und eines RK-Verfahrens unter Verwendung der 3/8-Regel als Quadraturformel (rechts).

0					0				
$1/2$	$1/2$				$1/3$	$1/3$			
$1/2$	0	$1/2$			$2/3$	$-1/3$	1		
1	0	0	1		1	1	-1	1	
	$1/6$	$1/3$	$1/3$	$1/6$		$1/8$	$3/8$	$3/8$	$1/8$

Konsistenz und Konvergenz Die Güte eines Einschrittverfahrens wird durch das sogenannte Restglied bzw. den Abschneidefehler $\mathbf{T}_h$ gemessen. Dieser gibt an, wie sich für einen einzelnen Integrationsschritt die Verfahrensfunktion $\Phi(t_j, \mathbf{Y}_j; h_j)$ von der theoretisch exakten Fortschreiterichtung unterscheidet.

Definition F.1.1 Zu einer aktuellen Näherung $(t_j, \mathbf{Y}_j) \in Q$ soll $z(t, t_j, \mathbf{Y}_j)$ die exakte Lösung der lokalen Anfangswertaufgabe bezeichnen:

$$\dot{z} = f(t, z(t)), \quad z(t_j) = \mathbf{Y}_j.$$

Damit definiert:

$$\mathbf{T}(t_j, \mathbf{Y}_j; h) := \frac{z(t_{j+h}, \mathbf{Y}_j) - \mathbf{Y}_j}{h} - \Phi(t_j, \mathbf{Y}_j; h)$$

den (lokalen) Abschnidefehler des Verfahrens. Ein Einschrittverfahren heißt dann konsistent mit der Ordnung $p \in \mathbb{N}$, wenn für alle Näherungen $(t_j, \mathbf{Y}_j)$ und hinreichend oft stetig differenzierbare rechte Seiten f eine Abschätzung der folgenden Form gilt:

$$\|\mathbf{T}(t_j, \mathbf{Y}_j; h)\|_\infty \leq C \cdot h^p, \quad \text{mit } C \cdot h^p \rightarrow 0 \quad \text{für } h \rightarrow 0,$$

wobei C eine von f und Q abhängige Konstante ist, s. Tabelle F–3 zur Definition der verwendeten Vektornorm.

Tab. F–3 Vektornormen

-
- Betragssummennorm (L_1 -Norm): $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$
 - Euklidnorm (L_2 -Norm): $\|\mathbf{x}\| = \sqrt{\sum_{i=1}^n x_i^2}$
 - Maximumnorm (L_∞ -Norm): $\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$
-

Der Konsistenzfehler wird manchmal auch durch: $\epsilon = z(t_j + h) - \mathbf{Y}_{j+1}$, definiert; ϵ hängt direkt mit dem hier verwendeten Fehler zusammen: $\epsilon = h_j \mathbf{T}(t_j, \mathbf{Y}_j; h_j)$, d.h. $\mathbf{T}$ ist der (absolute) Integrationsfehler pro Schrittweite.

Zur Bestimmung der Konsistenzordnung werden die Taylor-Reihenentwicklungen von $\Phi(t_j, \mathbf{Y}_j; h)$ bzgl. h (Entwicklungspunkt $h = 0$) und der exakten Lösung voneinander subtrahiert. Damit ergibt sich eine Konsistenzordnung von 1 für das Polygonzug-Verfahren, eine Ordnung von 2 für das modifizierte Euler und das Heun-Verfahren und eine Ordnung von 4 für die in Tabelle F–2 vorgestellten RK-Verfahren.

Die Konsistenz ist allein noch nicht hinreichend für die Konvergenz der Folge von Näherungen gegen die exakte Lösung der Anfangswertaufgabe. Bezogen auf einen festen Punkt t_b im Integrationsintervall besagt die Konvergenzaussage: $\mathbf{Y}(t_b; I_h) \rightarrow \mathbf{y}(t_b)$ ($\delta \rightarrow 0$), daß die Näherungswerte $\mathbf{Y}$ mit wachsendem Index auf verschiedenen Gittern die exakte Lösung $\mathbf{y}$ annähern sollen. Eine weitere notwendige Voraussetzung für die Konvergenz ist daher die Stabilität des Verfahrens. Sie bezeichnet die Eigenschaft des Einschrittverfahrens, für hinreichend kleine Rundungsfehler in den Anfangsbedingungen und in der sukzessiven Berechnung der Näherungslösungen an den Gitterpunkten auch nur kleine Abweichungen der berechneten Näherungslösung von der gesuchten exakten Lösung der Näherungsverfahren aufzuweisen, und zwar gleichmäßig mit der Schrittweite h .

Definition F.1.2 Ein Einschrittverfahren heißt stabil, wenn Konstanten S, r, H_0 existieren, so daß für je

zwei gestörte Lösungen $\mathbf{Y}(t_j; h_j), \tilde{\mathbf{Y}}(t_j; h_j)$:

$$\begin{aligned}\frac{1}{h_j}[\mathbf{Y}(t_{j+1}) - \mathbf{Y}(t_j)] &= \eta_j \\ \frac{1}{h_j}[\tilde{\mathbf{Y}}(t_{j+1}) - \tilde{\mathbf{Y}}(t_j)] &= \zeta_j,\end{aligned}$$

mit $j = 0, 1, \dots, m-1$ gilt:

$$\|\mathbf{Y}(t_j) - \tilde{\mathbf{Y}}(t_j)\|_\infty \leq S \left(\|\mathbf{Y}(t_0) - \tilde{\mathbf{Y}}(t_0)\|_1 + \|\eta_j - \zeta_j\|_\infty \right)$$

für alle Gitter I_h mit $h \leq h_0$ und solange $\|\mathbf{Y}(t_0) - \tilde{\mathbf{Y}}(t_0)\|_1 \leq r$ bzw. $\|\eta_j - \zeta_j\|_\infty \leq r$.

Mit anderen Worten folgt aus der Konsistenz und Stabilität die Konvergenz des Einschrittverfahrens.

Definition F.1.3 Das Einschrittverfahren sei konsistent mit der Ordnung $p > 0$ und außerdem stabil, z.B. aufgrund einer Lipschitz-Bedingung für Φ :

$$\|\Phi(t, \tilde{\mathbf{Y}}; h) - \Phi(t, \mathbf{Y}; h)\| = L_\Phi \|\tilde{\mathbf{Y}} - \mathbf{Y}\|.$$

Dann konvergieren die Näherungswerte $\mathbf{Y}(t; I_h)$ bei feiner werdendem Gitter $\delta \rightarrow 0$ gegen die exakte Lösung des Anfangswertproblems $\mathbf{y}(t)$. Für die Näherungen $\mathbf{Y}_m(t_b; I_h)$ im Endpunkt t_b gilt:

$$\|\mathbf{Y}(t_b; I_h) - \mathbf{y}(t_b)\| \leq \frac{C}{L_\Phi} (e^{L_\Phi(t_b - t_0)} - 1) \delta^p$$

Die Konvergenzordnung ist also p , d.h. sie ergibt sich aus der Konsistenzordnung.

Dabei wird die Differenz: $\epsilon(t; I_h) = \mathbf{y}(t) - \mathbf{Y}(t; I_h)$, als Gitterfehler bzw. lokaler Diskretisierungsfehler und die Maximumnorm $\|\epsilon(t; I_h)\|_\infty$ als globaler Diskretisierungsfehler bezeichnet.

Ist die numerische Lösung konvergent, gibt es aufgrund der Rundungsfehler bei einer vorgegebenen Maschinengenauigkeit eine Grenzgenauigkeit und eine dazugehörige optimale Schrittweite, bis zu der der Gesamtfehler des numerischen Verfahrens fällt und bei deren weiterer Verkleinerung wieder ansteigt.

Schrittweitensteuerung Bei der Schrittweitensteuerung geht es um die automatische Erzeugung eines Integrationsgitters I_h , das einerseits fein genug sein soll, um eine vorgegebene Genauigkeit der numerischen Integrationslösung zu garantieren, andererseits aber auch nicht feiner, um den numerischen Aufwand und den Einfluß von Rundungsfehlern möglichst gering zu halten. Meist wird durch den lokalen Prozeß eine nicht äquidistante Schrittweite generiert. Die Algorithmen zur Schrittweitensteuerung arbeiten mit einer Schätzung des lokalen Diskretisierungsfehlers, d.h. zu jedem Integrationsschritt $(t_j, \mathbf{Y}_j) \rightarrow (t_j + h, \mathbf{Y}_j + \Phi(t_j, \mathbf{Y}_j; h))$ mit einer aktuellen Schrittweite $h > 0$ wird ein numerisch berechenbarer Schätzwert $\tau_{est}(t_j, \mathbf{Y}_j; h)$ bestimmt:

$$\tau_{est}(t_j, \mathbf{Y}_j; h) \approx (1/h) [\mathbf{y}(t_{j+1}; t_j, \mathbf{Y}_j) - \mathbf{Y}_j - h\Phi(t_j, \mathbf{Y}_j; h)].$$

Für eine optimalen Schrittweite h_j^* wird eine vorgegebene Genauigkeitsanforderung, tol , gefordert:

$\|\tau(t_j, \mathbf{Y}_j; h_j^*)\| = \text{tol}$. Unter Berücksichtigung der Konsistenzordnung des Verfahrens folgt für die optimale Schrittweite:

$$h_j^* \approx \left(\frac{\text{tol}}{\|\tau_{est}\|} \right)^{1/p} h.$$

Eine effiziente Methode, den lokalen Diskretisierungsfehler zu schätzen, besteht in der Verwendung eingebetteter RK-Verfahren. Darunter wird ein Paar von RK-Verfahren mit gleicher Verfahrensmatrix und Knotenvektor aber unterschiedlichem Wichtungsvektor verstanden. Das eine Verfahren mit der Konsistenzordnung p ist das eigentliche Integrationsverfahren, das andere mit der Ordnung $p - 1$ oder $p - 1$ dient zur Schätzung des Diskretisierungsfehlers, wie z.B. beim RK7(8)-Verfahren (Fehlberg 1969) oder dem RK8(7)-Verfahren bzw. DOPRI8 (Prince und Dormand 1981).

F.1.3 Absolute (lineare) Stabilität und steife Systeme

Zur Herleitung der bisherigen Konvergenzaussagen, s. Definition F.1.3, wurden bisher Verhältnisse für feste Zeiten t_b und gegen Null gehende Schrittweiten h betrachtet. Dazu wurde die Stabilitätsdefinition in Abschnitt F.1.2 (inhärente Stabilität) für Einschrittverfahren bzw. die Nullstabilität (Wurzelbedingung) für Mehrschrittverfahren, s. unten, benötigt. In der Praxis werden jedoch feste Schrittweiten h verwendet, und evtl. eine Lösung für $t \gg t_0$ gesucht. Anstelle der Aussagen für genügend kleine Schrittweiten interessiert dann, daß die Näherungslösung für die gerade gewählte Schrittweite die exakte Lösung wenigstens qualitativ approximiert. Dafür wurden verschiedene Stabilitätskonzepte, u.a. auch für nichtlineare DGL-Systeme formuliert, wobei eine bzgl. dieser Konzepte instabile Näherungslösung die Konvergenzaussage nicht verletzt. Zum besseren Verständnis wird hier nur das Konzept der absoluten Stabilität anhand einer linearen Differentialgleichung erläutert. Um das Stabilitätsverhalten zu untersuchen, werden die Integrationsverfahren auf eine lineare Test- oder Modellgleichung angewendet, wie z.B. hier die skalare Gleichung:

$$\dot{y} = \lambda y(t), \quad y(t_0) = y_0. \quad (\text{F-8})$$

Bei negativen Realteil von λ gilt für die exakte Lösung: $|y(t_j)| \rightarrow 0$ falls $t_j \rightarrow \infty$. Nun stellt sich die Frage, ob sich auch die Näherungslösungen von Einschrittverfahren entsprechend verhalten, also $|Y_j(h)| \rightarrow 0$ für endliches $h > 0$ und $j \rightarrow \infty$.

Das explizite Euler-Cauchy-Verfahren, s. Gleichung (F-4), formuliert z.B. die Näherungslösung:

$$Y_{j+1}(h) = Y_j(h) + h\lambda Y_j(h) = (1 + h\lambda)Y_j(h).$$

Damit $|Y_j(h)| \rightarrow 0$, d.h. damit die Zahlenfolge $Y_j, Y_{j+1}, Y_{j+2}, \dots$ stetig abfällt, muß also gelten:

$$|1 + h\lambda| < 1 \quad \text{mit} \quad |1 + h\lambda| = \sqrt{(1 + h \operatorname{Re}(\lambda))^2 + (h \operatorname{Im}(\lambda))^2}.$$

Dies ist nie für ein rein imaginäres λ und nur dann für ein reelles $\lambda < 0$ erfüllt, wenn $h|\lambda| < 2$. Eine ähnliche Bedingung kann für das implizite Polygonzug-Verfahren formuliert werden:

$$\left| \frac{1}{1 - h\lambda} \right| = \frac{1}{\sqrt{(1 - h \operatorname{Re}(\lambda))^2 + (h \operatorname{Im}(\lambda))^2}} < 1,$$

die für alle $h > 0$ gültig ist, wenn $\operatorname{Re}(\lambda) < 0$ ist.

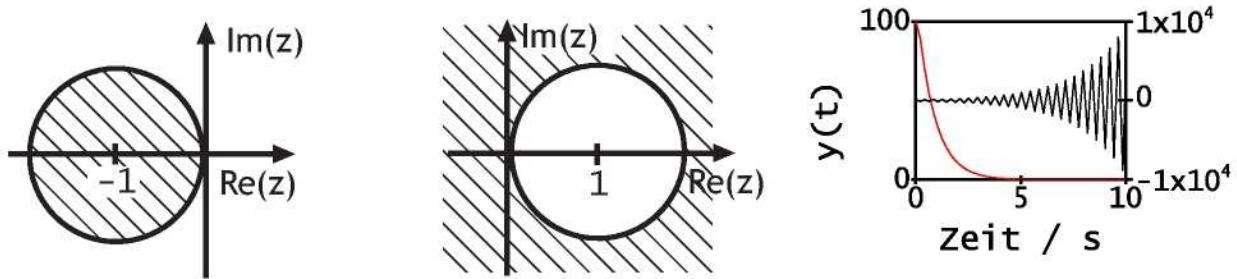


Abb. F-1 Stabilitätsgebiete des expliziten bzw. impliziten Euler-Cauchy-Verfahrens (links bzw. mitte); die exakte Lösung (rote Linie) der Testgleichung und die Approximation mit dem expliziten Euler-Cauchy-Verfahren (schwarze Linie) für $h=0,2$, $\lambda = -10$, $y_0=100$.

Allgemein ergibt sich bei Anwendung von Einschrittverfahren zur Approximation der exakten Lösung von Gleichung (F-8) der folgende Ansatz für die Näherungslösungen:

$$Y_{j+1;h} = g(z)Y_{j;h}, \quad z = h\lambda. \quad (\text{F-9})$$

$g(z)$ bezeichnet die Stabilitätsfunktion, die für das explizite Euler-Cauchy-Verfahren z.B. durch $g(z) = 1 + h\lambda$ definiert ist, s. oben. Mit der allgemein gültigen Stabilitätsfunktion läßt sich die Definition der absoluten linearen Stabilität ableiten.

Definition F.1.4 Ein Einschrittverfahren soll die Gleichung (F-9) für die Testgleichung: $\dot{y} = \lambda y$ erfüllen. Dann heißt die Teilmenge $\mathcal{G}$ der komplexen Ebene mit:

$$\mathcal{G} = \{z : |g(z)| < 1\}$$

das **Gebiet der absoluten Stabilität** des Einschrittverfahrens.

Das Einschrittverfahren heißt **A-stabil** (Dahlquist 1956), falls:

$$\mathcal{G} \supset \{z : \operatorname{Re}(\lambda) < 0\},$$

also, falls das Gebiet der absoluten Stabilität die linke Halbebene der komplexen Ebene umfaßt. Gilt zusätzlich:

$$g(z) \rightarrow 0 \quad \text{für} \quad \operatorname{Re}(z) \rightarrow -\infty$$

heißt das Verfahren **L-stabil**.

Die Stabilitätsgebiete des expliziten und des impliziten Polygonzug-Verfahrens sind in Abbildung F-1 aufgetragen. Das explizite Verfahren ist nicht A-stabil, das implizite Verfahren ist A- und L-stabil. Die Auswirkung dieser Feststellung, d.h. der Wahl einer falschen Schrittweite, auf die numerische Lösung des expliziten Polygonzug-Verfahrens ist in Abbildung F-1 (rechts) festgehalten. Da das Stabilitätsgebiet des impliziten Verfahrens auch die imaginäre Achse außer Null enthält wird die diskrete Lösung einer reinen Schwingungsgleichung mit $\operatorname{Re}(\lambda)$ gedämpft (künstliche Dissipation). Stabilitätseigenschaften in der Art, wie sie in diesem Abschnitt definiert wurden, werden für die praktische Charakterisierung eines steifen DGL-System verwendet: ein System wird steif genannt, wenn bei Verwendung von expliziten Verfahren die Schrittweite stärker durch Stabilitäts- als durch vorgegebene Genauigkeitsanforderungen eingeschränkt wird.

F.1.4 Mehrschrittverfahren

Verfahren Im Laufe der Zeit haben sich *zwei* Familien von linearen Mehrschrittverfahren in den Anwendungen durchgesetzt: die Adams-Verfahren für nichtsteife Probleme und die Rückwärts-Differentiationsformeln-Verfahren (backward differentiation formulas, BDF-Verfahren) für steife Systeme.

Die Adams-Verfahren ergeben sich aus der formalen Integration der Differentialgleichung in (F-1) über dem Intervall $[t_{j-k}, t_{j+l}]$, $k, l \geq 0$:

$$\mathbf{y}(t_{j+l}) - \mathbf{y}(t_{j-k}) = \int_{t_{j-k}}^{t_{j+l}} f(t, \mathbf{y}(t)) dt.$$

Dabei wird der Integrand durch ein Interpolationspolynom $P_s \in \Pi_s$ vom Maximalgrad s zu den Stützstellen: $(t_i, f(t_i, \mathbf{Y}_i)) =: (t_i, f_i)$, $i = j, j-1, \dots, j-s$, ersetzt. Wird eine für äquidistante Schrittweiten, $t_i = t_0 + ih$, $i = 0, 1, \dots$, vereinfachte Lagrange-Darstellung des Interpolationspolynoms gewählt und die Transformation $t = t_j + \tau h$, $dt = h d\tau$ durchgeführt, resultiert die allgemeine Form der Verfahren:

$$\begin{aligned} \mathbf{Y}_{j+l} &= \mathbf{Y}_{j-k} + h \sum_{i=0}^s b_{i,s} f_{j-i}, & f_{j-i} &:= f(t_{j-i}, \mathbf{Y}_{j-i}), \\ b_{i,s} &= \int_{-k}^l \prod_{\substack{p=0 \\ p \neq i}}^s \left(\frac{\tau + p}{-i + p} \right) d\tau, & i &= 0, 1, \dots, s. \end{aligned} \quad (\text{F-10})$$

Für $l = 1, k = 0$ ergibt sich z.B. das explizite Adams-Bashforth-Verfahren:

$$\begin{aligned} \mathbf{Y}_{j+1} &= \mathbf{Y}_j + h \sum_{i=0}^s b_{i,s} f_{j-i}, \\ b_{i,s} &= \int_0^1 \prod_{\substack{p=0 \\ p \neq i}}^s \left(\frac{\tau + p}{p - i} \right) d\tau := \beta_{i,s} / \gamma_s, \end{aligned} \quad (\text{F-11})$$

mit eindeutig bestimmten Koeffizienten $\beta_{i,s}, \gamma_s$ für jedes s . Für $s = 0$ ist z.B. $\beta_{i,s} = \gamma_{i,s} = 1$ d.h. das explizite Euler-Cauchy-Verfahren läßt sich als ein Spezialfall des Adams-Bashforth-Verfahren auffassen. Weitere Verfahren sind für andere Werte von k, l bekannt, s. Tabelle F-4.

Tab. F-4 Adams-Verfahren

k	l	Name	Art
0	1	Adams-Bashforth	Extrapolation explizites Verfahren
1	0	Adams-Moulton	Interpolation implizites Verfahren
1	1	Nyström	Extrapolation explizites Verfahren
2	0	Milne-Simpson	Interpolation implizites Verfahren

Die BDF-Verfahren werden gewonnen, indem das Interpolationspolynom $p_s \in \Pi_s$ zu den Stützstellen

$(t_{j+1-k}, \mathbf{Y}_{j+1-k}), k = 0, 1, \dots, s$ bestimmt:

$$p_s(t) = \sum_{k=0}^s \mathbf{Y}_{j+1-k} \ell_k \left(\frac{t_{j+1} - t}{h} \right),$$

$$\ell_k(\tau) = \prod_{m \neq k}^s \left(\frac{\tau - m}{k - m} \right) \quad (\text{äquidistanter Fall}),$$

und die Ableitung $\dot{\mathbf{y}}(t_{j+1})$ durch:

$$\dot{p}_s(t_{j+1}) = - \sum_{k=0}^s \frac{1}{h} \mathbf{Y}_{j+1-k} \dot{\ell}_k(0),$$

approximiert wird. Damit resultiert die allgemeine Form:

$$\sum_{k=0}^s c_{k,s} \mathbf{Y}_{j+1-k} = h f(t_{j+1}, \mathbf{Y}_{j+1}), \quad c_{k,s} = -\dot{\ell}_k(0). \quad (\text{F-12})$$

Die Koeffizienten für alle s sind eindeutig bestimmt, für Schrittzahlen $s > 6$ werden die Verfahren jedoch instabil.

Mehrschrittverfahren verlangen im Gegensatz zu den Einschrittverfahren beim Start eine Anlaufrechnung, mit der zunächst der Beginn der Datenkette $\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_s$ berechnet werden muß. Hierzu werden Einschrittverfahren oder Mehrschrittverfahren kleinerer Schrittzahl verwendet.

Um die Lösung der impliziten Gleichungen von Interpolationsverfahren zu vermeiden, werden explizite und implizite Adams-Verfahren zu den sogenannten Prädiktor-Korrektor-Verfahren gekoppelt. Dabei sagt das explizite Verfahren in jedem Integrationsschritt einen Wert $\mathbf{Y}_{j+k}^0$ voraus (P), an dem die Funktion f_{j+k}^0 evaluiert wird (E), mit der der alte iterierte Wert über das implizite Verfahren zu $\mathbf{Y}_{j+k}^1$ korrigiert wird (C). Die Schritte (E) und (C), d.h. das iterative Verfahren zur Lösung der impliziten Gleichung, wird dabei als Fixpunktiteration bezeichnet und beliebig oft, z.B. m Mal, durchgeführt. Je nachdem, ob dann die rechte Seite f im nächsten Integrationsschritt zunächst auf f_j^{m-1} oder f_j^m gesetzt wird, wird das Verfahren als P(EC)^m-Verfahren oder P(EC)^mE-Verfahren bezeichnet.

Konsistenz, Stabilität und Konvergenz Die allgemeine, in Gleichung (F-3) dargestellte Form eines linearen Mehrschrittverfahrens wird betrachtet, wobei:

$$\rho(z) := \sum_{k=0}^s \alpha_k z^k, \quad \sigma(z) := \sum_{k=0}^s \beta_k (z)^k,$$

die charakteristischen Polynome des Verfahrens bezeichnen. Weiterhin wird vorausgesetzt, daß das Mehrschrittverfahren mit äquidistanten Schrittweiten h durchgeführt wird und daß die Anfangswerte des Verfahrens $\mathbf{Y}_j = \mathbf{y}(t_j) + \epsilon_j, j = 0, 1, \dots, s-1$ die absoluten Fehler ϵ_k bei der Anlaufrechnung besitzen.

Definition F.1.5 Zu einer aktuellen Näherung $(t_j, \mathbf{Y}_j) \in Q$ soll $\mathbf{z}(t; t_j, \mathbf{Y}_j)$ die Lösung des lokalen AWP bezeichnen: $\dot{\mathbf{z}} = f(t, \mathbf{z}(t)), \mathbf{z}(t_j) = \mathbf{Y}_j$.

Zu hinreichend kleinem $h > 0$ heißt dann

$$\tau(t_j; \mathbf{Y}_j; h) := \frac{1}{h} \left[\sum_{k=0}^s \alpha_k \mathbf{z}(t_j + hk) - h \sum_{k=0}^s \beta_k \dot{\mathbf{z}}(t_j + hk) \right]$$

lokaler Diskretisierungsfehler des Verfahrens.

Das Mehrschrittverfahren heißt konsistent, falls für alle hinreichend oft stetig differenzierbaren rechten Seiten f und alle Näherungen $(t_j, \mathbf{Y}_j) \in Q$ eine Abschätzung der Form gilt:

$$\|\tau(t_j, \mathbf{Y}_j; h)\| \leq \phi(h), \quad \text{mit} \quad \phi(h) \rightarrow 0 \quad \text{für} \quad h \rightarrow 0.$$

Das Verfahren besitzt die Konsistenzordnung $p \in \mathbb{N}$, falls gilt:

$$\|\tau(t_j, \mathbf{Y}_j; h)\| \leq \phi(h) \quad \text{mit} \quad \phi(h) = O(h^p).$$

Die Konsistenz und Ordnungseigenschaften eines Mehrschrittverfahrens lassen sich mit Hilfe der charakteristischen Polynome überprüfen, s. (Deuffhard und Bornemann 2002, S. 338). Wie bei den Einschrittverfahren bestimmt neben den Ordnungseigenschaften die Stabilität die Konvergenz des Verfahrens.

Definition F.1.6 Ein lineares Mehrschrittverfahren heißt stabil (auch null-stabil), wenn alle Nullstellen z_j des charakteristischen Polynoms $\rho(z_j)$ in der Einheitskreisscheibe liegen ($|z_j| \leq 1$) und auf dem Rand ($|z_j| = 1$) nur einfache Nullstellen auftreten (Erfüllung der Dahlquistschen Wurzelbedingung). Liegen alle anderen Nullstellen außer der Nullstelle $z_1 = 1$ im Inneren des Einheitskreises, heißt das Verfahren stark bzw. strikt und ansonsten schwach stabil.

Die Konvergenz des Verfahrens wird in Analogie zu den Einschrittverfahren definiert.

Definition F.1.7 Vorgegeben ist ein beliebiges AWP: $\dot{\mathbf{y}}(t) = f(t, \mathbf{y}(t))$, $\mathbf{y}_0 = \mathbf{y}(t_0)$, mit einer (lokal) Lipschitz-stetigen rechten Seite f und einer auf dem Intervall $[t_0, t_b]$ erklärten exakten Lösung $\mathbf{y}(t)$.

Die Näherungslösungen $\mathbf{Y}(t_b; h_m)$ sollen mit der Gitterfolge $t_j^m := t_0 + jh_m$, $h_m := (t_b - t_0)/m$, $m \geq s$ und den Startnäherungen $\mathbf{Y}(t_j^m; h_m)$, $j = 0, 1, \dots, s-1$, wobei $\lim_{m \rightarrow \infty} |\mathbf{Y}(t_j; h_m) - \mathbf{y}(t_j)| = 0$, berechnet werden. Gilt dann stets:

$$\lim_{m \rightarrow \infty} |\mathbf{Y}(t_b; h_m) - \mathbf{y}(t_b)| = 0$$

heißt das Mehrschrittverfahren konvergent.

Gilt für alle hinreichend glatten rechten Seiten f und Startdaten mit $\lim_{m \rightarrow \infty} |\mathbf{Y}(t_j; h_m) - \mathbf{y}(t_j)| = O(h_m^p)$, daß auch im Endpunkt:

$$\lim_{m \rightarrow \infty} |\mathbf{Y}(t_b; h_m) - \mathbf{y}(t_b)| = O(h_m^p),$$

heißt das Mehrschrittverfahren konvergent von der Ordnung p .

Nach dem Satz von Dahlquist, ist ein lineares Mehrschrittverfahren konvergent, wenn es konsistent und stabil ist. Es hat darüber hinaus die Konvergenzordnung p , wenn die Konsistenzordnung p ist und die Startnäherungen die gleiche Approximationsgüte besitzen.

Adaptive Steuerung von Ordnung und Schrittweite Effiziente Verfahren zur Lösung von AWP benötigen eine adaptive (lokale) Bestimmung der Schrittweite. Dazu existieren im Wesentlichen zwei Ansätze, um die bisherige Voraussetzung der äquidistanten Schrittweiten bei der Konstruktion von Mehrschrittverfahren zu überwinden.

Der erste Ansatz verwendet anstatt der Lagrangeschen Darstellung des Interpolationspolynoms die Newtonsche Darstellung und kann dadurch mit völlig beliebigen Gittern arbeiten (Krogh 1974). Statt des sonst bei der Newton-Darstellung üblichen Rückwärtsdifferenzenquotienten bei äquidistantem Gitter, werden gitterabhängige dividierte Differenzen und gitterabhängige Koeffizienten verwendet. Damit ist zwar ein bequemer Ordnungswechsel möglich, aber eine relativ umfangreiche Nebenrechnung bei einem variablen Gitter erforderlich.

Ein anderer Vorschlag zur Behandlung variabler Gitter besteht darin, bei Schrittweitenänderung ein virtuelles äquidistantes Gitter mit der neuen Schrittweite in die Vergangenheit zu legen und die nötige Anfangsinformation zu interpolieren (Nordsieck 1962). Dazu werden die zur Approximation an der Stelle t_j notwendigen rückwärtigen Informationen über das Interpolationspolynom als Ableitungsinformation am letzten Integrationsknoten gespeichert und auf die Stelle t_j übertragen. Dadurch ist eine bequeme Änderung der Schrittweite bei konstanter Ordnung möglich, ein Wechsel der Ordnungen führt zu umfangreichen Nebenrechnungen.

Der hier verwendete DGL-Löser (Hindmarsh 1980) implementiert Adams- und BDF-Verfahren mit Ordnungs- und Schrittweitensteuerung in Nordsieck-Form.

F.2 Mathematische Optimierung

Die Beschreibungen in den folgenden Abschnitten folgen im Wesentlichen (Fletcher 1987; Kosmol 1993; Alt 2002; Sturm 2002).

F.2.1 Erklärung von Begriffen

Die Optimierung beschäftigt sich damit, die Minima und Maxima einer Funktion $F : R \rightarrow \mathbb{R}^k, k \in \mathbb{N}$:

$$F(x) = \begin{pmatrix} F_1(x) \\ \vdots \\ F_k(x) \end{pmatrix}, \quad x \in R, F_i : R \rightarrow \mathbb{R}, i = 1, \dots, k. \quad (\text{F-13})$$

über einer zulässigen Menge $R \subseteq \mathbb{R}^n, n \in \mathbb{N}$ zu finden. Jede Abbildung $F_i, i = 1, \dots, k$ heißt Zielfunktion. In der skalaren Optimierung ($k = 1$) werden Minima von F gesucht:

$$k = 1 : \quad \min_x \{F(x) \mid x \in R\}, \quad (\text{F-14})$$

während in der Mehrziel- bzw. Vektroptimierung ($k > 1$) konkurrierende Zielfunktionen abgeglichen werden müssen. Ist $R = \mathbb{R}^n$, so handelt es sich um eine kontinuierliche unrestringierte Optimierung (Optimierung ohne Nebenbedingungen). Sind $h_i : \mathbb{R}^n \rightarrow \mathbb{R}, i = 1, \dots, m + p$, Funktionen mit:

$$R = \{x \in \mathbb{R}^n \mid h_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = m + 1, \dots, p\} \subseteq \mathbb{R}^n,$$

so heißen die h_i Nebenbedingungen und es handelt sich um eine kontinuierliche restringierte Optimierung (Optimierung mit Nebenbedingungen). In beiden Fällen wird R ein kontinuierlicher zulässiger Bereich genannt.

Enthält $R \subset \mathbb{R}^n$ nur endlich viele oder unendlich abzählbar viele Elemente, so spricht man von diskreter Optimierung oder kombinatorischer Optimierung, und R wird diskreter zulässiger Bereich genannt.

Im Fall $R = R_1 \times R_2$, wobei $R_1 \subseteq \mathbb{R}^{n_1}$ ein kontinuierlicher zulässiger Bereich und $R_2 \subseteq \mathbb{R}^{n_2}$ ein diskreter zulässiger Bereich ist, spricht man von gemischt diskret-kontinuierlicher Optimierung (R heißt

gemischt diskret-kontinuierlicher zulässiger Bereich).

Ein skalares kontinuierliches restringiertes Optimierungsproblem mit linearen Nebenbedingungen h_i , wird lineare bzw. quadratische Optimierung genannt, wenn $F(\mathbf{x})$ linear bzw. quadratisch ist, ansonsten nichtlineare Optimierung. Für das Verständnis des nächsten Abschnitts werden noch die Definitionen des Gradienten-Vektors:

$$\nabla F(\mathbf{x}) := \begin{pmatrix} \frac{\partial F(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial F(\mathbf{x})}{\partial x_n} \end{pmatrix} \in \mathbb{R}^n \quad (\text{F-15})$$

der Hesse-Matrix benötigt:

$$\nabla^2 F(\mathbf{x}) := \begin{pmatrix} \frac{\partial^2 F(\mathbf{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 F(\mathbf{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F(\mathbf{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 F(\mathbf{x})}{\partial x_n \partial x_n} \end{pmatrix} \in \mathbb{R}^{n,n} \quad (\text{F-16})$$

F.2.2 Unrestringierte nichtlineare Optimierung

Für eine gegebene Zielfunktion $F: \mathbb{R}^n \rightarrow \mathbb{R}$ wird ein $\mathbf{x}^* \in \mathbb{R}^n$ mit $F(\mathbf{x}^*) \leq F(\mathbf{x}) \forall \mathbf{x} \in \mathbb{R}^n$ gesucht, s. Gleichung (F-14). Ein solches $\mathbf{x}^*$ heißt ein globaler Minimierer von F . Die Berechnung eines globalen Minimierers ist i.A. ein kompliziertes Problem, so daß meistens nur lokale Minimierer $\bar{\mathbf{x}} \in \mathbb{R}^n$ berechnet werden, für die gilt:

$$F(\bar{\mathbf{x}}) \leq F(\mathbf{x}) \quad \forall \mathbf{x} \in \mathbf{U}(\bar{\mathbf{x}}),$$

wobei $\mathbf{U}$ eine offene Umgebung von $\bar{\mathbf{x}}$ ist.

Das im letzten Absatz vorgestellte nichtlineare Optimierungsproblem kann in der Regel analytisch nicht gelöst werden. Daher erzeugen numerische Verfahren zur Berechnung der Minimierer ausgehend von einem gegebenen Startpunkt $\mathbf{x}_0 \in \mathbb{R}^n$ anhand von lokalen Informationen über die Zielfunktion (numerische Werte, Gradienten, Hessematrix, etc.) iterativ eine Reihe von Punkten $\mathbf{x}_j \in \mathbb{R}^n, j \in \mathbb{N}$.

Für die Konvergenz der Verfahren ist von Bedeutung, daß:

1. Die Zielfunktion in der Regel bestimmte mathematische Voraussetzungen erfüllt (Differenzierbarkeit, Verhalten im Unendlichen), die in Anwendungsfällen oft nicht überprüfbar oder gegeben sind;
2. Die Verfahrensvorschrift so konstruiert ist, daß unter gegebenen Voraussetzungen die Folge $\{\mathbf{x}_j\}_{j \in \mathbb{N}_0}$ gegen einen Minimierer $\bar{\mathbf{x}}$ konvergiert;

Konvergenzaussagen gelten in der Regel für unendliche Zahlenfolgen $\{\mathbf{x}_j\}_{j \in \mathbb{N}_0}$, während praktische numerische Verfahren bei Erreichen eines Abbruchkriteriums mit dem Punkt $\mathbf{x}_N \in \mathbb{R}^n$ als Approximation des Minimierers stoppen. Als Abbruchkriterien werden mit dem Computer prüfbare Optimalitätsbedingungen festgelegt, wobei zwischen notwendigen Bedingungen, die ein Minimierer erfüllen muß und hinreichenden Bedingungen, die garantieren, daß der betrachtete Punkt ein Minimierer ist, zu unterscheiden ist.

Besitzt ein Verfahren die notwendigen prinzipiellen Konvergenzeigenschaften, so wird zusätzlich eine möglichst hohe Konvergenzgeschwindigkeit erwünscht. Zur Charakterisierung der Konvergenzgeschwindigkeit wird die vom Algorithmus erzeugte Zahlenfolge $\{\mathbf{x}_j\}_{j \in \mathbb{N}_0}$ betrachtet, die gegen einen

Minimierer $\bar{\mathbf{x}}$ konvergiert. Die Zahlenfolge heißt:

- linear konvergent, falls es ein $\alpha \in \mathbb{R}$ mit $0 \leq \alpha < 1$ und ein $j_0 \in \mathbb{N}$ gibt, so daß:

$$\|\mathbf{x}_{j+1} - \bar{\mathbf{x}}\| \leq \alpha \|\mathbf{x}_j - \bar{\mathbf{x}}\| \quad \forall j \geq j_0.$$

- superlinear konvergent, falls es eine Nullfolge $\{\alpha_j\}_{j \in \mathbb{N}_0}$ ($\lim_{j \rightarrow \infty} \alpha_j = 0$) mit $\alpha_j > 0$ gibt, so daß:

$$\|\mathbf{x}_{j+1} - \bar{\mathbf{x}}\| \leq \alpha_j \|\mathbf{x}_j - \bar{\mathbf{x}}\| \quad \forall j \in \mathbb{N}_0.$$

- von p -ter Ordnung konvergent ($p > 1$), falls es ein $\alpha \geq 0$ und ein $j_0 \in \mathbb{N}$ gibt, so daß:

$$\|\mathbf{x}_{j+1} - \bar{\mathbf{x}}\| \leq \alpha \|\mathbf{x}_j - \bar{\mathbf{x}}\|^p \quad \forall j \geq j_0.$$

Die meisten bekannten numerischen Optimierungsalgorithmen sind Verfahren mit Liniensuche, d.h. der nächste Punkt der Zahlenfolge, $\{\mathbf{x}_j\}_{j \in \mathbb{N}_0}$, wird gemäß $\mathbf{x}_{j+1} = \mathbf{x}_j - \sigma \mathbf{s}_j$ berechnet, wobei σ bzw. $\mathbf{s}_j$ die mit der Liniensuche zu bestimmende Schrittweite, s. unten, bzw. Suchrichtung bezeichnet. Gradienten-Verfahren bzw. indirekte lokale Suchverfahren zur Ermittlung der Suchrichtung für differenzierbare Zielfunktionen sind z.B. die Methode des steilsten Abstiegs (die Suchrichtung ergibt sich aus der Richtung des entgegengesetzten Gradienten $\mathbf{s}_j := -\nabla F(\mathbf{x}_j)$, linear konvergentes Verhalten), die Newton-Methode, vgl. (Fletcher 1987), (zusätzliche Berechnung der invertierten Hesse-Matrix zur Gewinnung von Krümmungsinformationen: $\mathbf{s}_j := \nabla^2 F(\mathbf{x}_j)^{-1} \nabla F(\mathbf{x}_j)$, superlinear konvergentes Verhalten) oder Modifikationen des Newton-Verfahrens, wie die Quasi-Newton-Methoden (analog dem Newton-Verfahren, aber Approximationen der invertierten Hesse-Matrix werden verwendet, da sie nicht immer positiv definit und damit invertierbar ist), hierzu zählt z.B. die BFGS-Methode (Fletcher 1987; Shanno 1985; Broyden 1970), oder die Levenberg-Marquardt-Methode (Levenberg 1944; Marquardt 1963), bei der leicht veränderte Hesse-Matrizen benutzt werden. Daneben gibt es Liniensuchverfahren, die keine Ableitungen sondern nur die Funktionswerte der Zielfunktion zur Bestimmung der Suchrichtung verwenden. Die zyklische Koordinatenmethode ermittelt zur Berechnung des nächsten Punktes $\mathbf{x}_{j+1}$, d.h. in einem Iterationsschritt, optimale Schrittweiten, s. unten, sequentiell nacheinander in Suchrichtung aller Koordinatenachsen $\mathbf{e}_1, \dots, \mathbf{e}_n$ (explorative Suche). Die Methode zeigt ein relativ schlechtes Konvergenzverhalten und kann bei Anwendung auf nicht differenzierbare Funktionen evtl. in einem nicht-optimalen Punkt abbrechen. Um diese Nachteile zu beheben, führt z.B. das Verfahren von Hooke und Jeeves (1961) abwechselnd Iterationsschritte mit Suchrichtung entlang der Koordinatenachsen bzw. entlang der Richtung $\mathbf{x}_{j+1} - \mathbf{x}_j$ (Pattern Suche) durch. Das Rosenbrock-Verfahren (Rosenbrock 1960) rotiert zur Verbesserung der Koordinatenmethode bei jedem Iterationsschritt das Koordinatensystem in Richtung $\mathbf{x}_{j+1} - \mathbf{x}_j$ und führt wiederum eine explorative Suche aus.

Das Auffinden einer optimalen Schrittweite kann als ein eindimensionales Minimierungsproblem betrachtet werden, in das die vorher festgelegte Suchrichtung als Konstante eingeht, d.h. das Minimum der Zielfunktion wird entlang der durch die Suchrichtung vorgegebenen Linie gesucht (Liniensuche): $\min\{F(\mathbf{x}_j - \sigma \mathbf{s}_j) \mid 0 \leq \sigma \leq l\}$, l obere Grenze für die Schrittweite. Ableitungsfreie Liniensuchverfahren lassen sich in simultane bzw. sequentielle Suchverfahren einteilen. Bei simultanen Suchverfahren wird im Voraus entschieden, an welchen Stellen eines vorher festgelegten Unsicherheitsintervall für das Minimum die Zielfunktion ausgewertet werden soll (z.B. durch Zerlegung des Unsicherheitsintervalls in gleich große Teilintervalle bei der gleichförmigen Suche). Bei sequentiellen Verfahren wird auf Informationen aus vorherigen Funktionsauswertungen zurückgegriffen, um das Unsicherheitsintervall in möglichst wenigen Iterationen durch Wahl geeigneter Stellen einzugrenzen (z.B. bei der dichotomen Suche, Methode des goldenen Schnitts oder Fibonacci-Methode). Verfahren auf der Basis von Ableitungen setzen die einfache oder zweifache Differenzierbarkeit der Liniensuchfunktion voraus. Bei der

aufteilenden Suche wird der Wert der Zielfunktion in der Mitte des Unsicherheitsintervalls berechnet und anhand der Ableitung entschieden, ob das Minimum genau an der Stelle bzw. rechts oder links von ihr liegt, wobei in den letzten beiden Fällen das Unsicherheitsintervall pro Iterationsschritt halbiert wird. Die Newton-Methode, wie z.B. die Armijo Regel (Armijo 1966), berechnet auf der Basis der ersten und der zweiten Ableitung aufeinanderfolgende Punkte, die sich schrittweise dem Minimum annähern, anstatt das Unsicherheitsintervall iterativ zu reduzieren, wie die vorherigen Methoden.

Ziel von direkten Suchverfahren ist es, die aufwendigen Berechnungen der Liniensuchverfahren über eine heuristische Bestimmung der Schrittweite und der Suchrichtung zu umgehen. Dazu wird nur die Information über die Werte der Zielfunktion für bestimmte Auswertungspunkte $\mathbf{x}_j$ verwendet, um neue Auswertungspunkte $\mathbf{x}_{j+1}$ zu erzeugen. Daher sind sie für die Optimierung nicht-differenzierbarer, verrauschter Zielfunktionen geeignet. Zu den direkten Suchverfahren gehören z.B. die ursprünglichen Varianten der zyklischen Koordinatenmethode, des Hooke-Jeeves- und Rosenbrock-Verfahrens, die oben als Liniensuchverfahren erklärt wurden, die Simplex-Methode nach Nelder-Mead, s. nächsten Abschnitt, oder das Multidirektionale Suchverfahren (Torczon 1991).

Da alle bisher vorgestellten Verfahren bei jedem Schritt nur Punkte akzeptieren, die eine Verbesserung darstellen, d.h. der Funktionswert eines neuen Punktes ist kleiner als der Funktionswert des Ausgangspunktes, sind diese Algorithmen nur für die Suche nach lokalen Minimalstellen geeignet. Im Gegensatz dazu enthalten die sogenannten globalen Suchalgorithmen zusätzlich noch stochastische Elemente und können daher einfacher globale Optima bestimmen. Evolutionäre Algorithmen (Bäck und Schwefel 1993), wie die genetischen Algorithmen (Kinnebrock 1994), Evolutionsstrategien (Schwefel 1995) und die genetische Programmierung (Fogel et al. 1966), sind stochastische Suchverfahren, die von einer Population an individuellen Lösungen ausgehen. Gibt es ein globales Optimum, das sich in der Güte wesentlich von den anderen Optima unterscheidet, werden alle Lösungen ungefähr an der gleichen Stelle gefunden, bei lokalen Optima mit ähnlicher Güte sind die Lösungen verstreut. Die Lösungsfindung erfolgt generationsweise, indem pro Iterationsschritt auf der Basis von Rekombination und Mutation (nicht bei der genetischen Programmierung) neue Lösungen für die Auswertungspunkte aus den vorherigen Lösungen generiert werden. Infolge der Wahl des Partners bei der Rekombination und der Konkurrenz der „Kinder“- mit den „Eltern“- Lösungen erfolgt eine Selektion der jeweils besseren Lösungen. Bei stochastischen Verfahren, wie den Monte-Carlo-Methoden (Rinnooy Kan und Timmer 1989) oder dem Simulated-Annealing-Algorithmus (Kirkpatrick et al. 1983), wird in jedem Schritt der Auswertungspunkt zufällig geändert. Falls sich der Wert der Zielfunktion in einem Schritt verkleinert, wird der Auswertungspunkt auf jeden Fall akzeptiert, währenddessen ein Punkt, der eine Erhöhung des Wertes bewirkt, nur mit einer gewissen Wahrscheinlichkeit angenommen wird. Die Wahrscheinlichkeit ist bei Beginn der Optimierung groß und gegen Ende der Optimierung werden nur noch kleinere Änderungen des Auswertungspunktes zugelassen. Globale Suchalgorithmen stellen meist an den Suchraum und die Startbedingungen keine Anforderungen und können bei Optimierungsproblemen von hoher Dimension eingesetzt werden, sie erfordern jedoch einen relativ großen Simulationsaufwand.

F.2.3 Nelder-Mead-Simplexmethode und Subplex-Algorithmus

Das Nelder-Mead Simplexverfahren (Nelder und Mead 1965) wird in der Praxis häufig zur unrestringierten, nichtlinearen Optimierung verwendet, da die zu minimierende Zielfunktion keine Voraussetzungen, wie z.B. die Differenzierbarkeit, besitzen muß. Insbesondere werden zur Durchführung des Algorithmus lediglich Funktionswerte von $F(\mathbf{x})$ benötigt. Das Verfahren basiert auf der Konstruktion von Simplexes.

Definition F.2.1 Für $n + 1$ gegebene Vektoren $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^n$, wobei die Vektoren $\mathbf{x}_i - \mathbf{x}_0, i =$

$1, \dots, n$ linear unabhängig sind, bezeichnet die konvexe Hülle:

$$S = \left\{ \sum_{i=1}^n \lambda_i \mathbf{x}_i \mid \lambda_i \geq 0, i = 1, \dots, n, \sum_{i=1}^n \lambda_i = 1 \right\}$$

dieser $n + 1$ Punkte im Raum ein (n -dimensionales) Simplex (Polytop) mit den Ecken $\mathbf{x}_1, \dots, \mathbf{x}_n$.

Zum Start des Verfahrens wird ein Simplex S_0 vorgegeben. Jeder Iterationsschritt besteht dann aus den folgenden zwei Operationen:

- Bestimme für den aktuellen Simplex S_j die Ecke $\mathbf{x}_m$ mit dem größten Funktionswert:

$$F(\mathbf{x}_m) = \max \{F(\mathbf{x}_0), \dots, F(\mathbf{x}_n)\}$$

- Berechne einen Punkt $\mathbf{y}$ mit $F(\mathbf{y}) < F(\mathbf{x}_m)$ und ersetze $\mathbf{x}_m$ durch $\mathbf{x}_r$. Daraus resultiert der neue Simplex S_{j+1} mit den Ecken $\mathbf{x}_r$ und $\mathbf{x}_i, i = 0, \dots, n, i \neq m, r$.

Werden für $j = 0, \dots, n$ die Schwerpunkte der Ecken bzgl. $\mathbf{x}_j$ gegeben durch:

$$\mathbf{s}_j = \frac{1}{n} \sum_{\substack{i=0 \\ i \neq j}}^n \mathbf{x}_i,$$

können die Vorschriften zur Konstruktion des neuen Punktes wie folgt wiedergegeben werden:

- **Reflektion**

Ein neuer Punkt $\mathbf{x}_r$ wird durch Reflektion von Punkt $\mathbf{x}_m$ am Schwerpunkt $\mathbf{s}_m$ erhalten:

$$\mathbf{x}_r = \mathbf{s}_m + \gamma(\mathbf{x}_m - \mathbf{s}_m), \quad 0 < \gamma \leq 1.$$

- **Expansion**

Ein Punkt $\mathbf{x}_e$ wird durch die Verschiebung von $\mathbf{x}_m$ über $\mathbf{x}_r$ hinaus in Richtung $(\mathbf{x}_r - \mathbf{s}_m)$ erhalten:

$$\mathbf{x}_e = \mathbf{s}_m + \beta(\mathbf{x}_r - \mathbf{s}_m), \quad \beta > 1.$$

- **Kontraktion**

- **Innere partielle Kontraktion**

Ein Punkt $\mathbf{x}_c$ wird durch die Verschiebung von $\mathbf{x}_m$ zwischen $\mathbf{x}_m$ und $\mathbf{s}_m$ erhalten:

$$\mathbf{x}_c = \mathbf{s}_m + \alpha(\mathbf{x}_m - \mathbf{s}_m), \quad 0 < \alpha < 1.$$

- **Äußere partielle Kontraktion**

Der Punkt $\mathbf{x}_e$ wird durch die Verschiebung von $\mathbf{x}_m$ zwischen $\mathbf{s}_m$ und $\mathbf{x}_r$ erhalten:

$$\mathbf{x}_e = \mathbf{s}_m + \alpha(\mathbf{x}_r - \mathbf{s}_m), \quad 0 < \alpha < 1.$$

- **Totale Kontraktion**

Die Punkte $\mathbf{x}_i, i = 0, \dots, n \forall i \neq m$ werden durch die Mittelpunkte der Strecken von $\mathbf{x}_m$ nach $\mathbf{x}_i$ ersetzt oder nach (Rowan 1990) auf $\mathbf{x}_i$ verschoben:

$$\hat{\mathbf{x}}_i = \mathbf{x}_i + \frac{1}{2}(\mathbf{x}_j - \mathbf{x}_i) = \frac{1}{2}(\mathbf{x}_j + \mathbf{x}_i) \text{ bzw.}$$

$$\hat{\mathbf{x}}_i = \mathbf{x}_i + \delta(\mathbf{x}_j - \mathbf{x}_i), \quad 0 < \delta < 1.$$

Aufgrund der Konstruktionsprinzipien kann das Nelder-Mead Verfahren entwickelt werden:

1. Festlegen der Startparameter:

die Parameter $0 < \alpha < 1, \beta > 1, 0 < \gamma \leq 1$, und evtl. $0 < \delta < 1$, sowie der Startpunkt $\mathbf{x}_{0,0} \in \mathbb{R}^n$ werden gegeben; j , der Index für die Iterationsschritte, wird auf 0 gesetzt.

2. Festlegen der Eckpunkte des Startsimplex S_0 :

$$\mathbf{x}_{0,i} = \mathbf{x}_0 + \mathbf{s} \cdot \mathbf{e}_i, \quad i = 1, \dots, n,$$

wobei $\mathbf{e}_i$ den i -ten Einheitsvektor und $\mathbf{s}$ den Vektor mit den Schrittweiten zwischen $\mathbf{x}_0$ und den Ecken des Startsimplex in Richtung der jeweiligen Einheitsvektoren bezeichnet.

3. Bestimmung der Punkte $\mathbf{x}_{j,m}$, $\mathbf{x}_{j,t}$ und $\mathbf{x}_{j,s}$ gemäß:

$$F(\mathbf{x}_{j,m}) = \max\{F(\mathbf{x}_{j,0}), \dots, F(\mathbf{x}_{j,n})\},$$

$$F(\mathbf{x}_{j,t}) = \min\{F(\mathbf{x}_{j,0}), \dots, F(\mathbf{x}_{j,n})\},$$

$$F(\mathbf{x}_{j,s}) = \max\{F(\mathbf{x}_{j,i}) \mid i = 0, \dots, n, i \neq m\},$$

und des Schwerpunktes bzgl. $\mathbf{x}_{j,m}$:

$$\mathbf{s}_{j,m} = \frac{1}{n} \sum_{\substack{i=0 \\ i \neq m}}^n \mathbf{x}_{j,i}.$$

4. Berechnung des Punktes $\mathbf{x}_{j,r}$ durch Reflektion von $\mathbf{x}_{j,m}$ am Schwerpunkt $\mathbf{s}_{j,m}$:

$$\mathbf{x}_{j,r} = \mathbf{s}_{j,m} + \gamma(\mathbf{s}_{j,m} - \mathbf{x}_{j,m}).$$

5. (a) Falls $F(\mathbf{x}_{j,r}) < F(\mathbf{x}_{j,t})$, ist $\mathbf{x}_{j,r}$ ein neuer optimaler Punkt, daher wird geprüft, ob die Expansion in die dieselbe Richtung eine weitere Verbesserung ergibt:

$$\mathbf{x}_{j,e} = \mathbf{s}_{j,m} + \beta(\mathbf{x}_{j,r} - \mathbf{s}_{j,m}),$$

und $\mathbf{x}_{j,m}$ ersetzt:

$$\mathbf{x}_{j,m} = \begin{cases} \mathbf{x}_{j,e} & \text{falls } F(\mathbf{x}_{j,e}) < F(\mathbf{x}_{j,r}), \\ \mathbf{x}_{j,r} & \text{falls } F(\mathbf{x}_{j,r}) \leq F(\mathbf{x}_{j,e}) \end{cases}$$

(b) Falls $F(\mathbf{x}_{j,t}) \leq F(\mathbf{x}_{j,r}) \leq F(\mathbf{x}_{j,s})$, wird $\mathbf{x}_{j+1,m}$ auf $\mathbf{x}_{j,r}$ gesetzt.

(c) Falls $F(\mathbf{x}_{j,r}) > F(\mathbf{x}_{j,s})$, erfolgt eine weitere Fallunterscheidung:

i. Falls $F(\mathbf{x}_{j,r}) \geq F(\mathbf{x}_{j,m})$, ist es vermutlich ratsam, das Simplex nicht zu verlassen, daher erfolgt eine partielle innere Kontraktion:

$$\mathbf{x}_{j,c} = \mathbf{s}_{j,m} + \alpha(\mathbf{x}_{j,m} - \mathbf{s}_{j,m}).$$

ii. Falls $F(\mathbf{x}_{j,r}) < F(\mathbf{x}_{j,m})$, ist der Funktionswert an $\mathbf{x}_{j,r}$ zwar besser als an $\mathbf{x}_{j,m}$ aber schlechter als an allen anderen Ecken, so daß es evtl. ratsam ist, zwar außerhalb aber näher am vorherigen Simplex zu suchen. Daher wird eine partielle äußere Kontraktion durchgeführt:

$$\mathbf{x}_{j,c} = \mathbf{s}_{j,m} + \alpha(\mathbf{x}_{j,r} - \mathbf{s}_{j,m}).$$

Falls $F(\mathbf{x}_{j,c}) < F(\mathbf{x}_{j,m})$ wird $\mathbf{x}_{j+1,m} = \mathbf{x}_{j,c}$ gesetzt, ansonsten wird eine totale Kontraktion bzgl. des besten Punktes $\mathbf{x}_{j,t}$ durchgeführt:

$$\mathbf{x}_{j+1,i} = \frac{1}{2}(\mathbf{x}_{j,i} - \mathbf{x}_{j,t}) \quad \text{bzw.} \quad \mathbf{x}_{j+1,i} = \delta(\mathbf{x}_{j,i} - \mathbf{x}_{j,t}).$$

6. Der Index für die Iterationsschritte wird um eins erhöht: $j = j + 1$, und die Prozedur ab Schritt 3 wiederholt.

Als Abbruchskriterium wird in (Nelder und Mead 1965) vorgeschlagen, daß die Standardabweichung der Funktionswerte an den Simplexecken kleiner als eine vorgegebene Toleranzgrenze sein soll, d.h.:

$$\frac{1}{n+1} \left(\sum_{i=0}^n (F(\mathbf{x}_{j,i}) - \bar{F}_j)^2 \right)^{1/2} < tol,$$

wobei:

$$\bar{F}_j = \frac{1}{n+1} \sum_{i=0}^n F(\mathbf{x}_{j,i})$$

den Mittelwert der Funktionswerte an den Simplexecken bezeichnet.

Der Nelder-Mead-Algorithmus hat folgende Eigenschaften:

- Im Vergleich zu anderen Methoden zeigt der Algorithmus bei verrauschten Funktionen ein gutes Konvergenzverhalten.
- Die Geschwindigkeit des Algorithmus ist für Probleme von niedriger Dimension ($n \leq 5$) groß, für höhere Dimensionen jedoch drastisch erniedrigt (Box 1966).
- Es gibt keine allgemeinen Konvergenzaussagen und keine Garantie, daß das Verfahren tatsächlich ein (zumindest lokales) Minimum von $F(\mathbf{x})$ liefert. Für spezielle Funktionen und niedrige Raumdimensionen werden Konvergenzresultate erzielt (Lagarias et al. 1998). Jedoch sind diese Resultate nicht allgemein gültig, da in (McKinnon 1998) Beispiele konstruiert werden, für die das Verfahren gegen nicht-stationäre Punkte konvergiert^{6.1}.
- Der Algorithmus erfordert einen Speicherbedarf von $\mathcal{O}(n^2)$, weil ein Simplex $n+1$ Ecken hat und jede Ecke einen Vektor der Dimension n darstellt.

Um die positiven Eigenschaften des Nelder-Mead-Algorithmus beizubehalten und die Effizienz des Verfahrens für Optimierungsaufgaben mit hoher Dimension zu verbessern, wurde die Subplex-Methode entwickelt (Rowan 1990), die den Raum $\mathbb{R}^n$ in Subräume zerlegt, d.h. jeweils einige Komponenten des Vektors mit den Auswertungspunkten werden zur Optimierung selektiert. Sie ist den im letzten Abschnitt erläuterten Liniensuchverfahren ähnlich, die in jedem Schritt optimale Suchrichtungen bestimmen und die Schrittweite über Liniensuche optimieren. Denn das Subplex-Verfahren ermittelt in jedem Schritt verbesserte Subräume und sucht die Minima in den Subräumen mit dem Nelder-Mead-Algorithmus (NMS: Nelder-Mead Simplex). Die Subräume werden so bestimmt, daß die Komponenten des Vektors $\mathbf{x}_j$, deren Änderungen die stärkste Auswirkung auf den Wert der Zielfunktion haben, in einem Subraum liegen. Am Ende des Iterationsschrittes setzt sich der neue Vektor $\mathbf{x}_{j+1}$ aus den in jedem Subraum gefundenen minimalen Komponenten für $\mathbf{x}_j$ zusammen und es gilt: $\Delta \mathbf{x}_{j+1} := \mathbf{x}_{j+1} - \mathbf{x}_j$.

Damit kann der Algorithmus wie folgt zusammengefaßt werden.

^{6.1} Es gibt eine konvergente Modifikation des Verfahrens, vgl. (Price et al. 2002a), die jedoch hier nicht verwendet wurde.

Subplex Methode**Eingabe**

n : Dimension des Vektors mit den Auswertungspunkten
 $F(\mathbf{x})$: Zielfunktion
 $\mathbf{x}_0$: Startvektor der Auswertungspunkte
 $\mathbf{s}_0$: Vektor mit den Schrittweiten in die n Koordinatenrichtungen
 α : Koeffizient für die NMS-Kontraktion $0 < \alpha < 1$
 β : Koeffizient für die NMS-Expansion $\beta > 1$
 γ : Koeffizient für die NMS-Reflektion $0 < \gamma \leq 1$
 δ : Koeffizient für die totale Kontraktion (NMS) $0 < \delta < 1$
 ψ : NMS-Reduktionskoeffizient $0 < \psi < 1$
 ω : Koeffizient für die Reduktion der Schrittweite $0 < \omega < 1$
 $n_{\min}$: minimale Dimension eines Subraums
 $n_{\max}$: maximale Dimension eines Subraums
 $1 \leq n_{\min} \leq n_{\max} \leq n \cup n_{\min} \lfloor n/n_{\max} \rfloor \leq n$

Ausgabe

$\mathbf{x}_n$: Lage des berechneten Minimums
 $F(\mathbf{x}_n)$: Berechnetes Minimum der Zielfunktion

begin

Initialisierung: Setze $\Delta \mathbf{x}_0$ gleich $\mathbf{s}_0$

while Abbruchsbedingung nicht erfüllt **do**

begin

Berechne die Subräume

Berechne den Vektor mit den Schrittweiten $\mathbf{s}_j$

for jeden Subraum **do**

begin

Berechne das Startsimplex

Bestimme das Minimum im Subraum mit dem Nelder-Mead-Algorithmus

end

Berechne das Abbruchkriterium

end

end

Die aktuelle Zahl n_{subs} und die Dimension der Subräume $n_{s_i}, i = 1, \dots, n_{\text{subs}}$ müssen folgenden Beziehungen genügen:

$$\sum_{i=1}^{n_{\text{subs}}} n_{s_i} = n, \quad n_{\min} \leq n_{s_i} \leq n_{\max}.$$

Zur Bestimmung der Subräume werden zunächst die Komponenten von $\Delta \mathbf{x}_j$ entsprechend der absoluten Größe ihres Wertes sortiert:

$$(\Delta x_1, \dots, \Delta x_n)_j^T \Rightarrow (\Delta x_{p_1}, \dots, \Delta x_{p_n})_j^T$$

wobei $|\Delta x_{p_{i+1}}| \geq |\Delta x_{p_i}| \quad \forall i = 1, \dots, n-1.$

Die Dimension des ersten Subraums ist der Wert von k , für den die folgenden (diskreten) Funktionen ein Maximum einnehmen:

$$n_{s1} = \begin{cases} \arg \max_k \left\{ \frac{\|(\Delta x_{p_1}, \dots, \Delta x_{p_k})_j^T\|_1}{k} - \frac{\|(\Delta x_{p_{k+1}}, \dots, \Delta x_{p_n})_j^T\|_1}{n-k} \right\} & \text{falls } k < n, \\ \arg \max_k \left\{ \frac{\|(\Delta x_{p_1}, \dots, \Delta x_{p_n})_j^T\|_1}{n} \right\} & \text{falls } k = n. \end{cases}$$

wobei k die Nebenbedingungen:

$$n_{smin} \leq k \leq n_{smax} \quad \text{und} \quad n_{smin}[(n-k)/n_{smax}] \leq n-k,$$

erfüllen muß.

Damit werden die $(x_{p_1})_j, \dots, (x_{p_k})_j$ Komponenten des Vektors $\mathbf{x}_j$ im ersten Subraum optimiert, und die Suche des nächsten Subraums erfolgt mit dem $(n - n_{s1})$ -Vektor analog zur Suche des ersten Subraums.

Anhand der Berechnung des Schrittweitenvektors wird festgestellt, um welchen Betrag die einzelnen Komponenten in den Subräumen verändert werden müssen, um die Ecken der Start-Simplizes für die innere Schleife zu erhalten. Der Vektor wird zu Beginn des Subplex-Algorithmus initialisiert und ergibt sich dann bei jeder Iteration j der äußeren Schleife gemäß:

$$\mathbf{s}_{j+1} = \begin{cases} \min(\max(\|\Delta \mathbf{x}_j\|_1 / \|\mathbf{s}_j\|_1, \omega), 1/\omega) & \text{falls } n_{subs} > 1, \\ \psi \cdot \mathbf{s}_j & \text{falls } n_{subs} = 1. \end{cases}$$

ψ ist der vorher festgelegte Faktor, um den die Größe des letzten NMS, d.h. $\|(\mathbf{x}_m - \mathbf{x}_t)\|$, in Bezug auf den vorletzten NMS reduziert werden muß, damit die Nelder-Mead-Suche im Subraum abgebrochen werden kann. Für $n_{subs} = 1$ werden die Schrittweiten zur Erstellung des Simplex für den nächsten Iterationsschritt im gleichen Ausmaß wie bei der letzten vorherigen Nelder-Mead-Iteration reduziert, damit die nächste Nelder-Mead-Suche mit dem vorherigen Stand fortgesetzt werden kann. Für $n_{subs} > 1$ gilt: wenn die Lage des Punktes $\mathbf{x}_{j+1}$ sich in Bezug auf $\mathbf{x}_j$ nur wenig verändert hat, wird die Schrittweite verringert, ansonsten vergrößert. Der Koeffizient ω gibt dabei das Ausmaß vor, mit dem die Schrittweite reduziert werden soll. Bei großen Werten für ω konvergiert der Algorithmus schnell in ein evtl. lokales Minimum, ansonsten erfolgt eine längere Suche. Das Vorzeichen der i -ten Komponente der Schrittweite ergibt sich folgendermaßen:

$$(s_i)_{j+1} = \begin{cases} \text{signum}((\Delta x_i)_j) \cdot |(s_i)_{j+1}| & \text{falls } \Delta x_i \neq 0, \\ -(s_i)_{j+1} & \text{falls } \Delta x_i = 0, \end{cases}$$

d.h. die Werte der einzelnen Komponenten $(x_{p_i})_j$ können bei der Erzeugung der Start-Simplizes erhöht oder erniedrigt werden.

Die Abbruchsbedingung für die innere Schleife wurde bereits erwähnt, die Bedingung für die äußere Schleife lautet:

$$\frac{\max(\|\Delta \mathbf{x}\|_\infty, \|\mathbf{s}\|_\infty \cdot \psi)}{\max(\|\mathbf{x}\|_\infty, 1)} \leq tol$$

Anhand der kurzen Darstellung des Verfahrens können die wesentlichen Eigenschaften des Verfahrens im Vergleich zu dem Nelder-Mead-Verfahren abgeleitet werden:

- Die Subplex-Methode verallgemeinert den Nelder-Mead-Algorithmus und die zyklische Koordinatenmethode :
 - wenn $nsmin = nsmax = n$, gibt es nur einen $\mathbb{R}^n$ Raum. Wird weiterhin der Wert von ψ entsprechend klein gewählt, bleibt der Algorithmus in der inneren Schleife, d.h. das Nelder-Mead-Suchverfahren wird durchgeführt.
 - wenn $nsmin = nsmax = 1$, ist jeder Subraum eindimensional, und das Verfahren entspricht der zyklischen Koordinatenmethode.
- Da das Verfahren auf dem Nelder-Mead-Algorithmus basiert, wird es sich auch für vertauschte Zielfunktionen eignen.
- In Bezug auf das Nelder-Mead-Verfahren benötigt der Algorithmus weniger Speicherbedarf: $\mathcal{O}(n) + \mathcal{O}(nsmax^2)$, da niemals ein Simplex mit mehr als $nsmax$ Ecken gespeichert werden muß.
- Die Geschwindigkeit des Algorithmus ist für mittlere bis große Dimensionen n durch die Aufteilung in Subräume erhöht ($n = 1, \dots, 32$, getestet in (Rowan 1990)).
- Zu verbesserten Eigenschaften bzgl. Konvergenz des Verfahrens im Vergleich zu dem Nelder-Mead-Algorithmus können nicht unbedingt Aussagen getroffen werden.

Anhang G Ergänzungen zu den Stoffwechselmodellen

G.1 Parameterwerte der Modelle zur Beschreibung der AAA-Synthese

Die Parameterwerte für die vorgestellten fünf Modelle zur Beschreibung der Synthese aromatischer Aminosäuren in zwei *E.coli*-Mutanten, s. Abschnitt 5.2, sind in den Tabellen G–1 und G–2 gegeben.

Tab. G–1 Parameterwerte der Modelle 1 und 2 aus Abschnitt 5.2

Modell 1 zur Beschreibung von Experiment MU1				Modell 1 zur Beschreibung von Experiment MU2			
Reaktions- kinetik	Parameter	Wert	Dimension	Reaktions- kinetik	Parameter	Wert	Dimension
$v_{pep \rightarrow out}$	$k_{pep \rightarrow out}$	$1 \cdot 10^{-6}$	1 / s	$v_{pep \rightarrow out}$	$k_{pep \rightarrow out}$	1,46	1 / s
v_{aroF}	k_{aroF}	4,44	1 / (mM · s)	v_{aroF}	k_{aroF}	11,39	1 / (mM · s)
v_{aroA}	k_{aroA}	0,13	1 / (mM · s)	v_{aroA}	k_{aroA}	1,54	1 / (mM · s)
v_{aroB}	$V_{max,aroB}$	181,78	mM / s	v_{aroB}	$V_{max,aroB}$	3606,83	mM / s
	$K_{m,aroB}$	$2 \cdot 10^{-6}$	mM		$K_{m,aroB}$	0,03	mM
	$K_{i,aroB}$	$9 \cdot 10^{-3}$	mM		$K_{i,aroB}$	$6 \cdot 10^{-4}$	mM
$v_{dah \rightarrow out}$	$k_{dah \rightarrow out}$	$4 \cdot 10^{-12}$	1 / s	$v_{dah \rightarrow out}$	$k_{dah \rightarrow out}$	0,84	1 / s
v_{aroD}	k_{aroD}	10,01	1 / s	v_{aroD}	k_{aroD}	0,29	1 / s
$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out}$	1,01	1 / s	$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out}$	0,15	1 / s
$v_{aroE,hin}$	k_{aroE}	0,10	1 / s	$v_{aroE,hin}$	k_{aroE}	0,06	1 / s
$v_{aroL,hin}$	$k_{aroL,hin}$	0,18	1 / (mM · s)	$v_{aroL,hin}$	$k_{aroL,hin}$	0,25	1 / (mM · s)
$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out}$	0,13	1 / s	$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out}$	0,82	1 / s
v_{Glyk}	k_{Glyk}	8,86	1 / s	v_{Glyk}	k_{Glyk}	29,19	1 / s

Modell 2 zur Beschreibung von Experiment MU2							
Reaktions- kinetik	Parameter	Wert	Dimension	Reaktions- kinetik	Parameter	Wert	Dimension
$v_{pep \rightarrow out}$	$k_{pep \rightarrow out}$	1,43	1 / s	$v_{aroE,hin}$	$K_{eq,aroE}$	3,00	mM
v_{aroF}	k_{aroF}	0,10	1 / (mM · s)		$K_{m,aroE,NADP}$	0,90	mM
v_{aroA}	$V_{max,aroA}$	2,00	mM / s		$K_{m,aroE,DHS}$	0,05	mM
	$K_{m,aroA,S3P}$	$3 \cdot 10^{-3}$	mM	$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out}$	0,16	1 / s
	$K_{m,aroA,PEP}$	0,02	mM	$v_{aroE,rück}$	$k_{aroE,rück}$	$1 \cdot 10^{-3}$	1 / (mM · s)
v_{aroB}	$V_{max,aroB}$	3606,82	mM / s	$v_{aroL,hin}$	$V_{max,aroL}$	10,00	mM / s
	$K_{m,aroB}$	0,03	mM		$K_{m,aroL,Shik}$	5,00	mM
	$K_{i,aroB}$	$6 \cdot 10^{-4}$	mM		$K_{m,aroL,ATP}$	0,16	mM
$v_{dah \rightarrow out}$	$k_{dah \rightarrow out}$	0,08	1 / s	$v_{aroL,rück}$	$k_{aroL,rück}$	$1 \cdot 10^{-4}$	1 / (mM · s)
v_{aroD}	$V_{max,aroD}$	0,28	mM / s	$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out}$	0,78	1 / s
	$K_{m,aroD}$	0,01	mM	v_{Glyk}	k_{Glyk}	29,19	1 / s
$v_{aroE,hin}$	$V_{max,aroE}$	0,50	mM / s				

Die Fehler der Parameter sind in allen Modellen wesentlich höher als die Parameterwerte selbst ($\Delta p/p = 10^2$ bis 10^9). Diesbezüglich gelten auch für diese noch relativ einfachen Modelle die gleichen Aussagen, welche im Abschnitt 6.2.6 für das komplexere Stoffwechselmodell 10 zur Beschreibung des *E.coli*-Wildstamms getroffen wurden. D.h. die Parameterzahl der Modell hätte über nachfolgende Sensitivitätsanalysen zur Eliminierung und Zusammenfassung der Parameter, s. Abschnitt 6.2.4 und Abschnitt 6.2.5, reduziert werden müssen, um die Parameter mit einer höheren Genauigkeit angeben zu können.

Weiterhin ist zu bedenken, daß bei den Experimenten MU1 und MU2 die Meßfehler-Kovarianzmatrix

aus einer Größtfehlerabschätzung, s. Abschnitt 4.2, hervorgeht, während bei Experiment WT1 ein wesentlich geringerer Fehler für die Meßdaten angesetzt wurde, s. Abschnitt 6.2.6^{7.1}. Es zeigt sich also an den großen Parameterfehlern für die Modelle zur Beschreibung der Experimente MU1 und MU2 deutlich, daß die Qualität der Meßergebnisse die Genauigkeit der Modellaussagen stark beeinflusst. Zur Verringerung der Meßunsicherheiten sollten die Experimente daher mehrfach unter den gleichen Versuchsbedingungen wiederholt werden, s. auch die Diskussion in Abschnitt 4.2.

Tab. G–2 Parameterwerte der Modelle 3, 4 und 5 aus Abschnitt 5.2

Modell 3 zur Beschreibung von Experiment MU2				Modell 4 zur Beschreibung von Experiment MU2			
Reaktions- kinetik	Parameter	Wert	Dimension	Reaktions- kinetik	Parameter	Wert	Dimension
$v_{c5p \rightarrow dhq}$	k_1	$8 \cdot 10^4$	1 / (mM · s)	$v_{c5p \rightarrow dhq}$	k_1	$8 \cdot 10^4$	1 / (mM · s)
v_{aroA}	k_{aroA}	0,40	1 / (mM · s)	v_{aroD}	k_{aroD}	12,22	1 / (mM · s)
v_{aroD}	k_{aroD}	14,11	1 / s	$v_{dhq \rightarrow out}$	$k_{dhq \rightarrow out}$	$1 \cdot 10^{-5}$	1 / s
$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out}$	7,16	1 / s	$v_{aroE, hin}$	$V_{max, aroE}$	93,00	mM / s
v_{aroE}	k_{aroE}	0,13	1 / s		$K_{eq, aroE}$	$3 \cdot 10^4$	mM
$v_{aroL, hin}$	$k_{aroL, hin}$	0,40	1 / (mM · s)		$K_{m, aroE, DHS}$	32,82	mM
$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out}$	1,67	1 / s		$K_{m, aroE, NADP}$	0,92	mM
				$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out}$	0,73	1 / s
				$v_{aroL, hin}$	$k_{aroL, hin}$	643,35	1 / (mM · s)
				$v_{aroL, rück}$	$k_{aroL, rück}$	$6 \cdot 10^4$	1 / (mM · s)
				v_{aroA}	k_{aroA}	0,11	1 / (mM · s)
				$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out}$	10,09	1 / (mM · s)

Modell 5 zur Beschreibung von Experiment MU2							
Reaktions- kinetik	Parameter	Wert	Dimension	Reaktions- kinetik	Parameter	Wert	Dimension
$v_{c5p \rightarrow dhq}$	k_1	363,07	1 / (mM · s)	$v_{aroE, hin}$	$K_{m, aroE, NADP}$	$3 \cdot 10^{-4}$	mM
v_{aroD}	k_{aroD}	2,95	1 / s	$v_{dhs \rightarrow out}$	$k_{dhs \rightarrow out}$	0,44	1 / s
$v_{dhq \rightarrow out}$	$k_{dhq \rightarrow out}$	$1 \cdot 10^{-5}$	1 / s	$v_{aroL, hin}$	$V_{max, aroL}$	64,04	mM / s
$v_{aroE, hin}$	$V_{max, aroE}$	99,99	mM / s		$K_{m, aroL, Shik}$	318,05	mM
	$K_{eq, aroE}$	$3 \cdot 10^4$	mM	v_{aroA}	k_{aroA}	4,99	1 / (mM · s)
	$K_{m, aroE, DHS}$	1,65	mM	$v_{s3p \rightarrow out}$	$k_{s3p \rightarrow out}$	3,75	1 / s

G.2 Zusammenfassung der Resultate für den *E.coli*-Wildstamm

Insgesamt 130 Modelle wurden erstellt, um die Stoffwechselvorgänge in der Glykolyse und dem PPP während eines Substrat-Puls-Experimentes mit einem *Escherichia coli* K12 Stamm zu beschreiben. Die Modelle lassen sich nach dem jeweiligen Reaktionsnetzwerk, das zugrundegelegt wurde, in 6 verschiedene Modell-Topologien einteilen, s. Abb. G–1 und G–2:

- 0.a: Ausgangsstruktur für alle Modelle zur Beschreibung der Glucose-Aufnahmerate und der anschließenden Aufteilung des Flusses in einen Fluß in die Glykolyse und einen in den PPP;

^{7.1} Aufgrund der mangelnden Informationen über dieses Experiment konnte keine exakte Fehlerrechnung durchgeführt werden. Statt dessen wurde die in Gleichung (6–30) mit $f_{rel} = ug = 0,05 \text{ mol l}^{-1}$ für alle Metabolite und Zeitpunkte verwendet. Mit dieser Gleichung wurden relative Meßfehler zwischen ca. 2% und 7% für die intrazellulären Metabolitkonzentrationen abgeschätzt, während aus der Größtfehlerabschätzung für die Meßdaten aus den Experimenten MU1 und MU2 zwei- bis dreifach höhere Meßunsicherheiten berechnet wurden.

- I: Struktur für eine sehr vereinfachte Beschreibung der Vorgänge in der Glykolyse und im PPP;
- II: Struktur mit detaillierterer Beschreibung des PPP und der Glykolyse in Richtung des Zitronensäurezyklus hin. Für die isobaren und damit experimentell nicht aufzutrennenden Pentose-5-Phosphate wurde ein Summenpool (C5P) angenommen;
- III: Aus der Struktur II generierte Struktur, die jedoch immer die Spaltung von FBP in GAP und DHAP berücksichtigt. Wahlweise wurden zusätzliche Reaktionen für eine alternative Glucose-Aufnahme, den Entner-Doudoroff-Weg und verschiedene Biomasseabflüsse postuliert;
- IV: Aus Struktur III generierte Struktur, die eine noch detailliertere Beschreibung des PPP enthält;
- V: Aus Struktur III generierte Struktur, die durch Formulierung unterschiedlicher Bruttoreaktionen innerhalb des Pentose-Phosphat-Weges versucht, den Metaboliten Erythrose-4-Phosphat aus den Modellen zu eliminieren;

Innerhalb jeder dieser Strukturen wurden durch Zuordnung verschiedener kinetischer Gleichungen zu den einzelnen Reaktionsschritten, s. Abschnitt G.3, mehrere Modelle erstellt.

Es zeigte sich, daß die Strukturen I und II die Glykolyse und den Pentose-Phosphat-Weg zu stark vereinfachen, da hier nur die Konzentrationsverläufe einiger, der experimentell bestimmten Metabolite wiedergegeben werden konnten.

Mit mehreren Modellen der Topologie III (Modelle 34, 36, 37, 53, 55, 56, 74, 76, s. Abschnitt G.3) gelang die Anpassung aller bei der Modellierung berücksichtigten Metabolite, wobei die experimentellen Gleichgewichtskonzentrationen relativ gut und die Konzentrationen nach dem Puls innerhalb von Größenordnungen beschrieben werden konnten. Nicht ganz exakt war die Wiedergabe der Veränderungen der F6P- und der 6PG-Konzentration unmittelbar nach dem Glucose-Puls, sowie des Schwingens der C5-Körper durch die Modelle, s. in Abb. G-3 einige Ergebnisse des Modells 37 stellvertretend für die Modelle der Struktur III.

Um hauptsächlich die Anpassung der Metabolite des PPP zu verbessern, wurde Struktur IV mit einer detaillierten Beschreibung des PPP entwickelt. Hierbei lagen jedoch weder für die E4P noch für die S7P-Konzentration experimentelle Ergebnisse vor. Keines der Modelle dieser Struktur konnte die experimentellen Daten annähern, da hier offenbar ein über-parametrisiertes Modell vorlag und daraus Schwierigkeiten bzgl. der Konvergenz des Optimierungsalgorithmus resultierten. Infolge der Erkenntnisse für die Modelle aus Struktur IV wurde zur Vermeidung der Über-Parametrisierung der Modelle aus Struktur III die Struktur V generiert, um auch den nicht gemessenen Metaboliten E4P zu eliminieren. Die Modelle 65 und 66 aus Struktur V, s. Abschnitt G.3, konnten die experimentellen Trajektorien aller Metabolite wiedergeben. Die Dimension der Modelle verringerte sich durch die Eliminierung von E4P von 10 DGL und ca. 100 Parametern auf 9 DGL und ca. 90 Parameter. Die Güte der Anpassung ist vergleichbar zu den beschriebenen Modellen aus Struktur III, der zeitliche Verlauf der Pyruvat-Konzentration konnte in Modell 66 sogar noch etwas besser beschrieben werden.

Im PPP ließ sich jedoch keine Verbesserung der Anpassungsgüte feststellen, s. Abb. G-4 für einige Ergebnisse des Modells 66. Bei dem Vergleich der simulierten Flüsse, s. Abb. G-3 und Abb. G-4 jeweils unten, zeigt sich, daß die Glucose-Aufnahmeraten und der Fluß in die Glykolyse bei beiden Modellen ähnlich sind, die Flüsse innerhalb des PPP und in den AAA sich jedoch unterscheiden. Charakteristisch für Modell 65 und 66 von Struktur V ist der hohe Fluß in den PPP und von dort aus in den AAA.

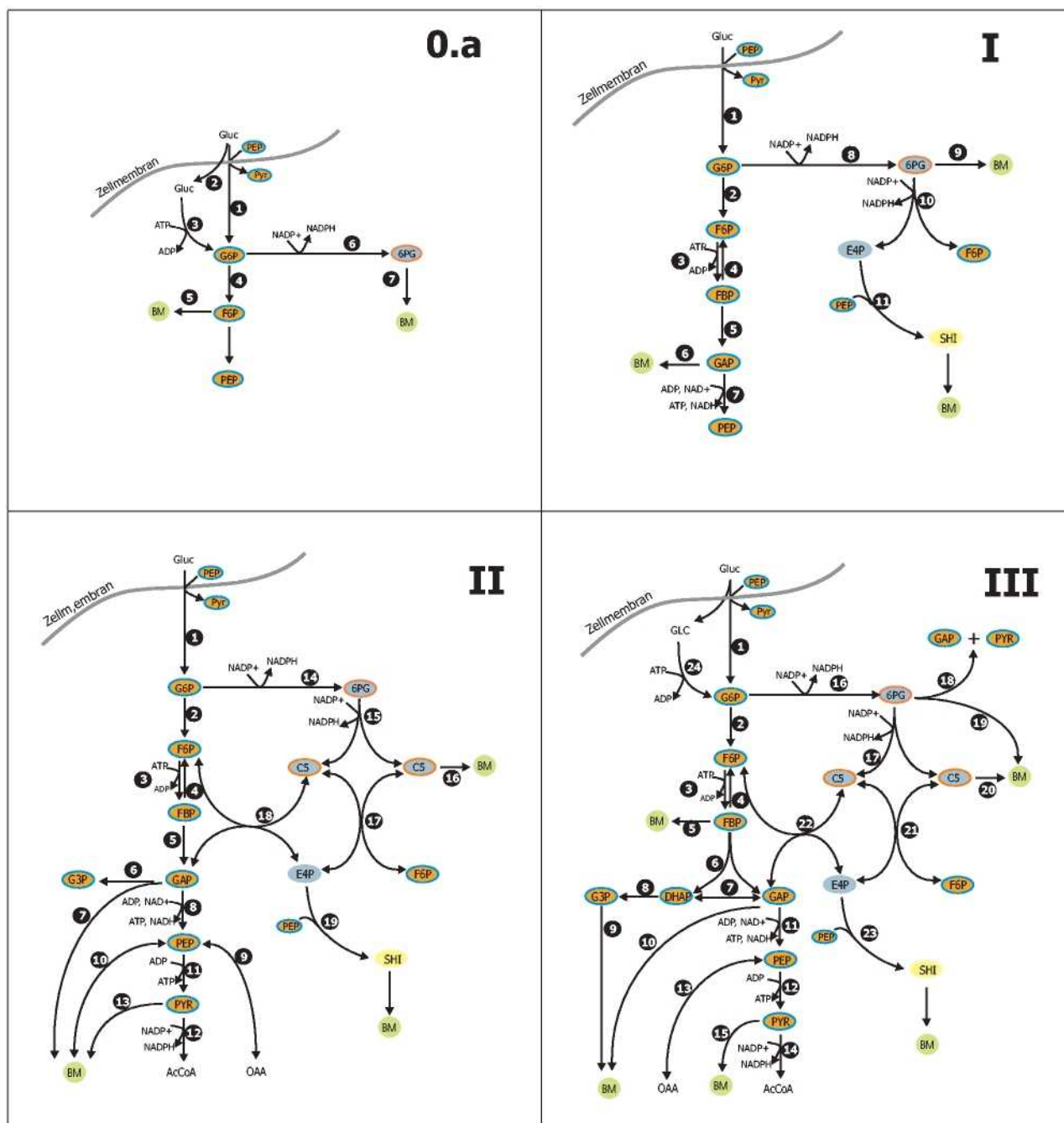


Abb. G–1 Teil 1 der Modelltopologien; umrandete Ellipsen: Metabolite, für die Meßdaten vorhanden sind und DGL erstellt wurden; Ellipsen ohne Rand: Metabolite, für die keine Daten vorhanden sind und DGL erstellt wurden; einfacher Text: Metabolite, deren zeitabhängiger Verlauf aus den Daten eingelesen wurde; Kreise: Annahme eines Abfluß in die Biosynthese (nicht-bilanziert)

Somit ist es also offenbar möglich, mit verschiedenen Modellstrukturen unterschiedliche Flußverteilungen bei in etwa gleicher Wiedergabequalität der Konzentrationen zu simulieren.

Insofern wurde mit weiteren Modellen untersucht, ob die Anpassung der Meßdaten auch bei Annahme eines vom Verzweigungspunkt G6P ausgehenden größeren Fluß in die Glykolyse als in den PPP gelingt. Dies wäre nämlich in biologischer Hinsicht einer der signifikantesten Unterschiede bzgl. der Aussagen, die überhaupt mit Modellen für die Stoffwechselabschnitte der Glykolyse und des PPP erzielt werden können.

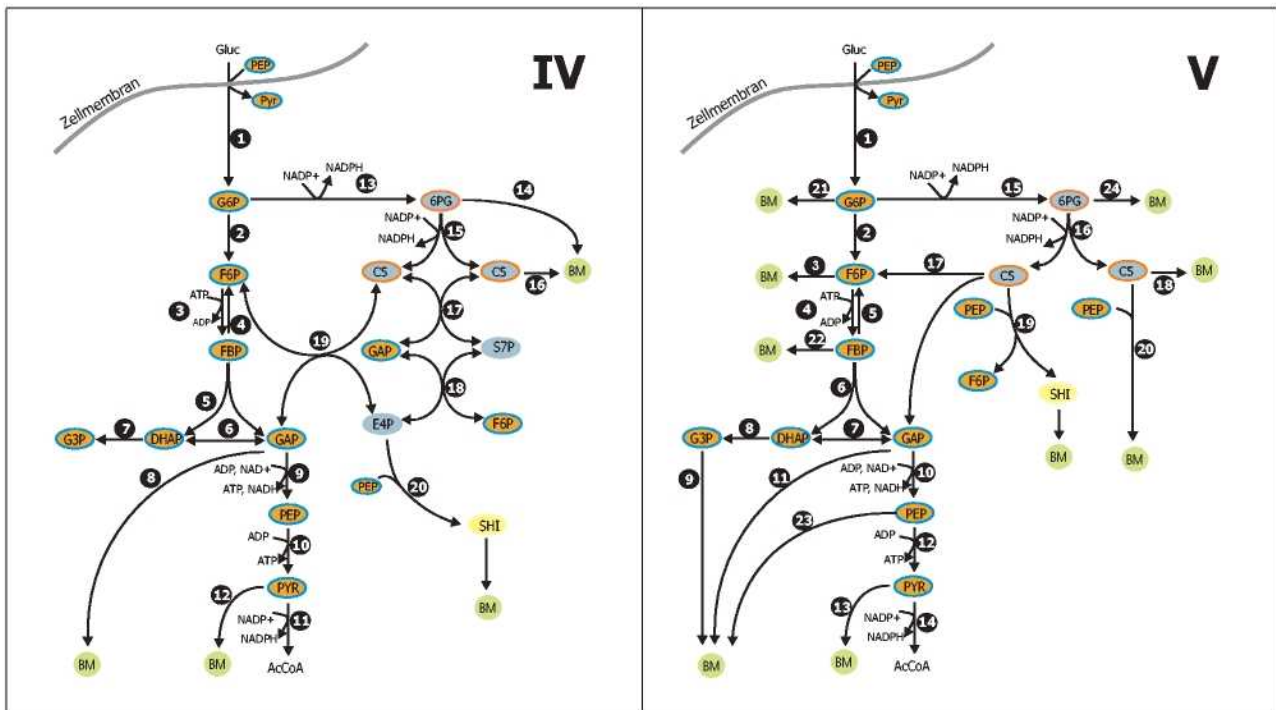


Abb. G–2 Teil 2 der Modelltopologien; umrandete Ellipsen: Metabolite, für die Meßdaten vorhanden sind und DGL erstellt wurden; Ellipsen ohne Rand: Metabolite, für die keine Daten vorhanden sind und DGL erstellt wurden; einfacher Text: Metabolite, deren zeitabhängiger Verlauf aus den Daten eingelesen wurde; Kreise: Annahme eines Abfluß in die Biosynthese (nicht-bilanziert)

Zu diesem Zweck wurden 62 weitere Modelle auf der Basis von Struktur V konstruiert. Mit den Modellen 69, 70 und 75, kann nun auch der Metabolit G3P wiedergegeben werden, die Metabolite des PPP werden vor allem in der Zeit vor dem Puls besser als bei den vorherigen Modellen beschrieben, und auch die GAP- bzw. die DHAP-Anpassung gelingt besser. Innerhalb eines Modells können jedoch entweder die Metabolite PEP und FBP oder die Metabolite Pyruvat und G6P+F6P jeweils weniger exakt als vorher wiedergegeben werden. In Abb. G–5 sind die Simulationsergebnisse für die Konzentrationen und Flüsse anhand von Modell 69 dargestellt. Mit diesem Modell ist die Anpassungsqualität der experimentellen Trajektorien von PEP und FBP also geringer als bei den vorigen Modellen, s. Abb. G–3 und G–4.

Im Vergleich zu den vorherigen Modellen kehrt sich das Verhältnis zwischen dem Fluß in die Glykolyse und dem Fluß in den PPP in Bezug auf die vorherigen Modelle um.

Es wird jedoch immer noch die Annahme eines relativ großen Abflusses aus dem PPP, diesmal in die Biosynthese der Nukleotide, benötigt, um eine gute Anpassungsgüte für die experimentelle Trajektorie von 6PG zu erreichen. Als Vorteil der Modelle wird bewertet, daß die Aldolase nicht mehr den Stofffluß durch die Glykolyse limitiert, wie bei den vorherigen Modellen. Denn dieser Umstand ist nicht aus der Literatur bekannt.

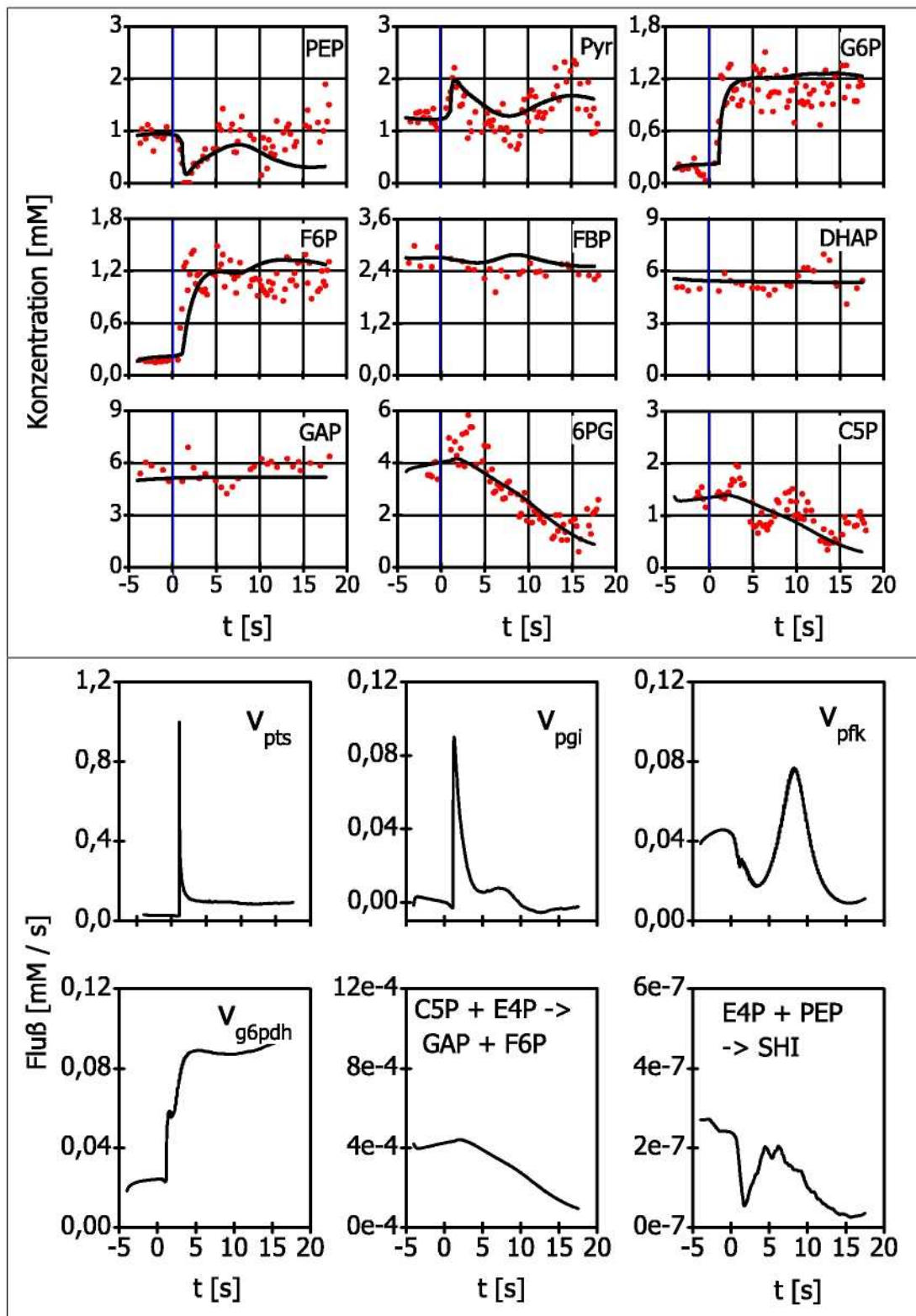


Abb. G-3 Ausgewählte Ergebnisse von Modell 37; oben simulierte Konzentrationsverläufe; unten simulierte Flüsse

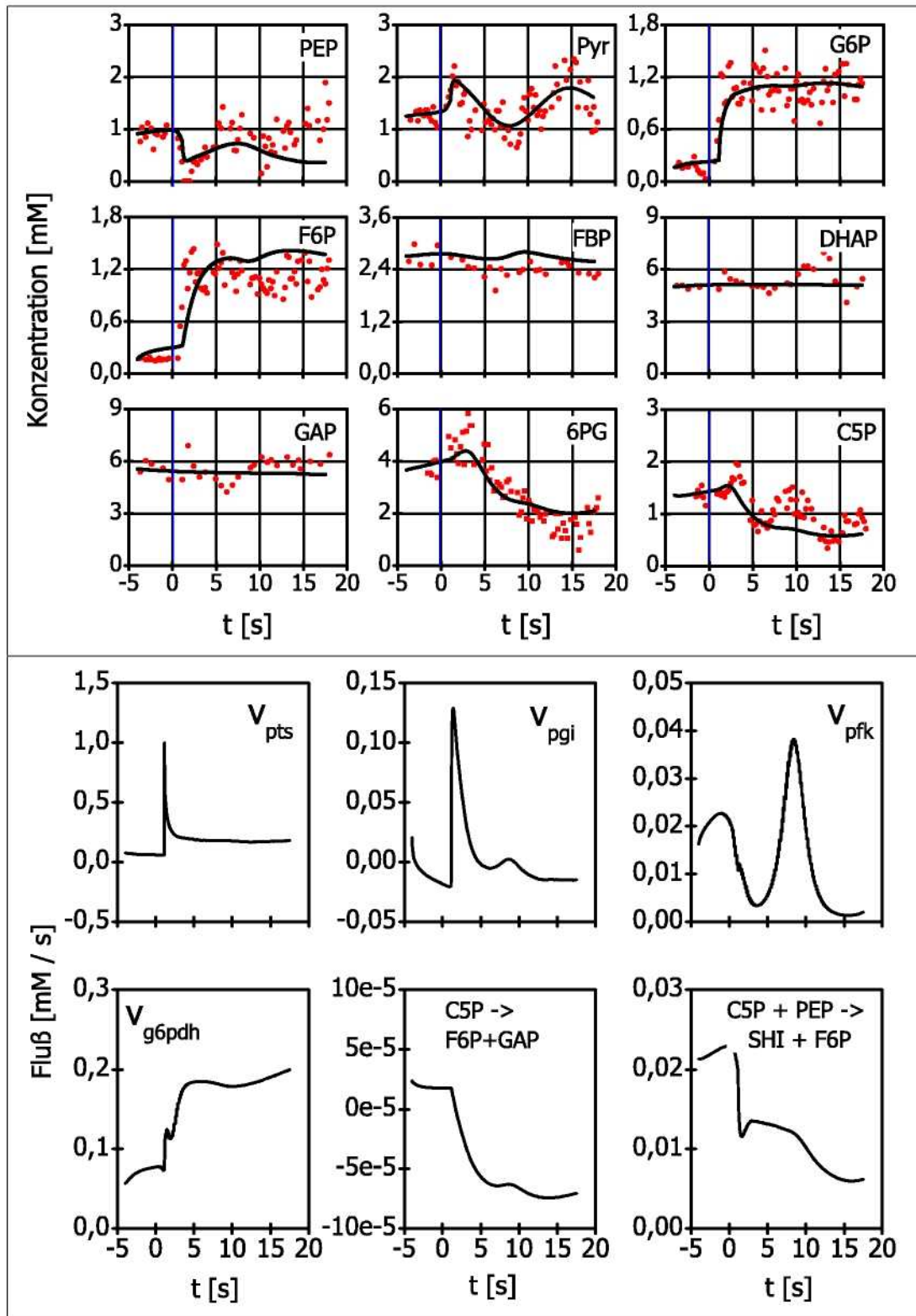


Abb. G-4 Ausgewählte Ergebnisse von Modell 66; oben simulierte Konzentrationsverläufe; unten simulierte Flüsse

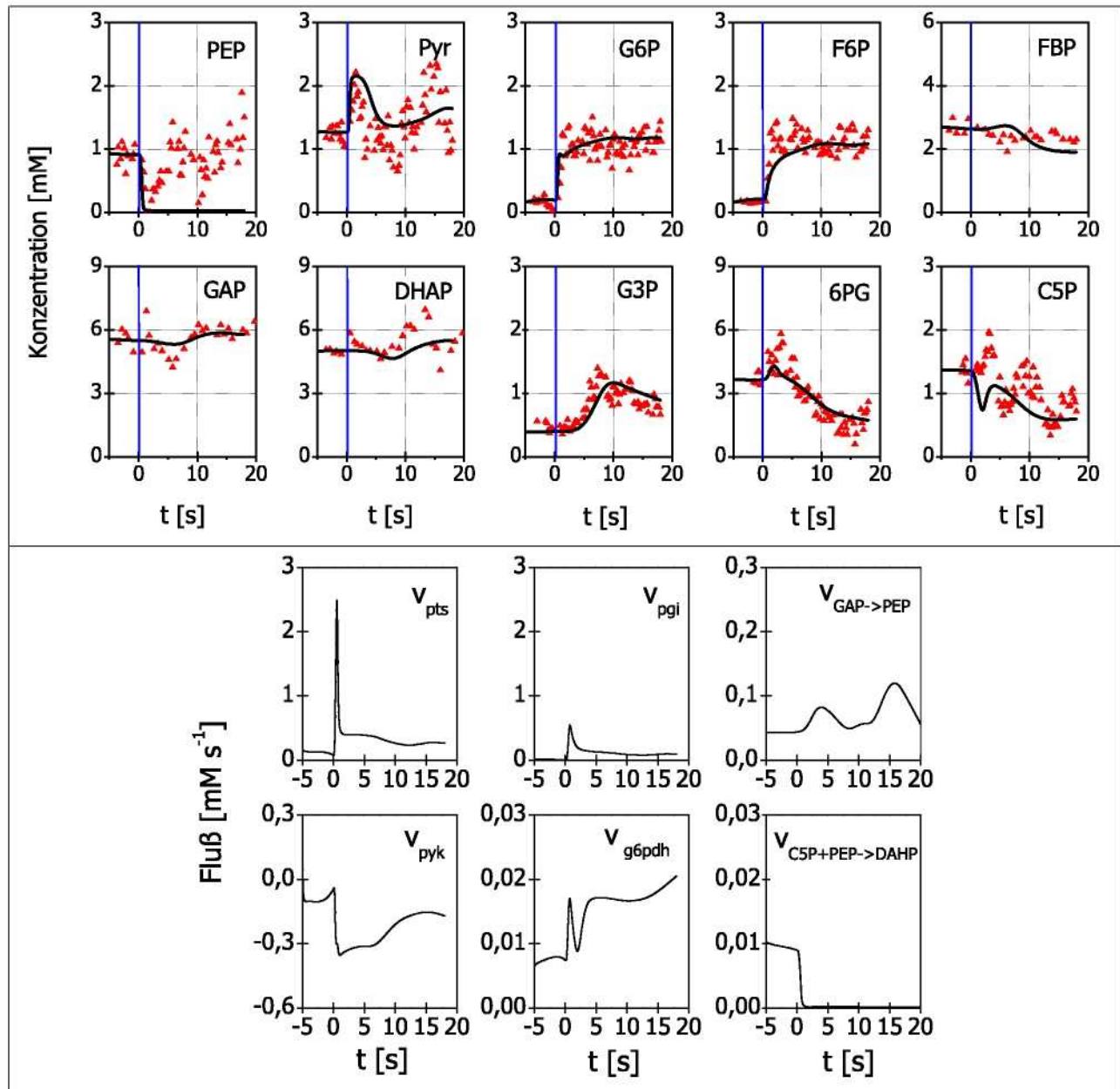


Abb. G-5 Ausgewählte Ergebnisse von Modell 69; oben simulierte Konzentrationsverläufe; unten simulierte Flüsse

G.3 Auflistung der Modelle für den *E.coli*-Wildstamm

In diesem Abschnitt werden die 76 der 130 Modelle vorgestellt, mit denen die Wiedergabe von wenigstens einigen der experimentellen Trajektorien gelang. Dazu wurden den Reaktionsschritten in den Strukturen 0.a-V Nummern zugewiesen, s. Abb. G–1 und G–2, und alle verwendeten Kinetiken in den Tab. G–3 aufgelistet. Aus den Tabellen G–4 bis G–7 können dann für die einzelnen Reaktionsnummern die Kinetiken des jeweiligen Modells entnommen werden. Z.B. wurde der Reaktion 1 (Glucose-Aufnahme) der Struktur 0.a im Modell 72 die Kinetik 7, d.h. eine Doppelsubstrat-Kinetik, zugewiesen, s. Tab. G–4, linke Tabelle, erste Zeile.

Abschließend werden in Tab. G–8 die Modellnummern der im Hauptteil beschriebenen, zwölf besten Modelle den verwendeten Nummern im Anhang zugeordnet, so daß diese Modelle im Detail nachvollzogen werden können.

Tab. G–3 Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1

1. *Geschwindigkeitsgesetz erster Ordnung für eine unimolekulare irreversible Reaktion*

$$v = k \cdot C \quad (\text{G-1})$$

2. *Irreversible Michaelis-Menten Gleichung*

$$v = \frac{V_{\max} \cdot C}{K_m + C} \quad (\text{G-2})$$

3. *Irreversible Michaelis-Menten Gleichung mit unkompetitiver Inhibition*

$$v = \frac{V_{\max} \cdot C}{K_m + C \cdot \left(1 + \frac{C}{K_i}\right)} \quad (\text{G-3})$$

4. *Irreversible Michaelis-Menten Gleichung mit gemischter Inhibition*

$$v = \frac{V_{\max} \cdot C}{K_m \cdot \left(1 + \frac{I}{K_i}\right) + C \cdot \left(1 + \frac{I}{\alpha K_i}\right)}, \quad \text{a) } I = \text{FBP} \quad (\text{G-4})$$

5. *Irreversible Michaelis-Menten Gleichung mit nicht-kompetitiver Inhibition*

$$v = \frac{V_{\max} \cdot C}{K_m \cdot \left(1 + \frac{I}{K_i}\right) + C \cdot \left(1 + \frac{I}{K_i}\right)}, \quad \text{a) } I = \text{G6P} \quad (\text{G-5})$$

6. *Irreversible Michaelis-Menten Gleichung mit Substrat-Aktivierung*

$$v = \frac{V_{\max} \left(\frac{C}{K_{a1}}\right)^2}{1 + \frac{C}{K_{a1}} + \left(\frac{C}{K_{a1}}\right)^2 + \frac{C}{K_{a2}}}, \quad \text{a) } C = \text{PEP} \quad (\text{G-6})$$

7. *Doppelsubstrat-Kinetik*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{(K_{m,1} + C_1) \cdot (K_{m,2} + C_2)} \quad (\text{G-7})$$

a) $C_1 = \text{GAP}, C_2 = \text{NAD}$; b) $C_1 = \text{C5P}, C_2 = \text{ATP}$; c) $C_1 = \text{F6P}, C_2 = \text{ATP}$;

8. *Doppelsubstrat-Kinetik mit einem multiplikativen Inhibitor-Term*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{(K_{m,1} + C_1) \cdot (K_{m,2} + C_2)} \cdot \frac{1}{(K_i + I)^n} \quad (\text{G-8})$$

a) $I = \text{G6P}$, b) $I = \text{Pyr}$, c) $I = \text{NADP}$, d) $I = \text{AcCoA}$

Tab. G–3 Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT19. *Doppelsubstrat-Kinetik mit drei multiplikativen Inhibitor-Termen*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{(K_{m,1} + C_1) \cdot (K_{m,2} + C_2)} \cdot \frac{1}{(K_{i,1} + I_1)^{n_1}} \cdot \frac{1}{(K_{i,2} + I_2)^{n_2}} \cdot \frac{1}{(K_{i,3} + I_3)^{n_3}} \quad (\text{G-9})$$

a) $I_1 = \text{F6P}$, $I_2 = \text{GAP}$, $I_3 = \text{AMP}$;10. *Trisubstrat-Kinetik*

$$v = \frac{V_{\max} \cdot C_1}{K_{m,1} + C_1} \cdot \frac{C_2}{K_{m,2} + C_2} \cdot \frac{C_3}{K_{m,3} + C_3} \quad (\text{G-10})$$

11. *Reversible Michaelis-Menten Gleichung*

$$v = \frac{V_f \cdot \left(\frac{C_1}{K_{m,1}} \right) - V_r \cdot \left(\frac{C_2}{K_{m,2}} \right)}{1 + \left(\frac{C_1}{K_{m,1}} \right) + \left(\frac{C_2}{K_{m,2}} \right)} \quad (\text{G-11})$$

12. *Vereinfachte Uni-Bi-Kinetik*

$$v = \frac{V_f \cdot \left(\frac{C_1}{K_{m,1}} \right) - V_r \cdot \left(\frac{C_2 \cdot C_3}{K_{m,2}} \right)}{1 + \left(\frac{C_1}{K_{m,1}} \right) + \left(\frac{C_2 \cdot C_3}{K_{m,2}} \right)} \quad (\text{G-12})$$

13. *Reversible Michaelis-Menten Gleichung mit einem multiplikativen Inhibitor-Term*

$$v = \frac{V_f \cdot \left(\frac{C_1}{K_{m,1}} \right) - V_r \cdot \left(\frac{C_2}{K_{m,2}} \right)}{1 + \left(\frac{C_1}{K_{m,1}} \right) + \left(\frac{C_2}{K_{m,2}} \right)} \cdot \left(\frac{1}{K_i + I} \right)^n, \quad \text{a) } I = \text{GAP} \quad (\text{G-13})$$

14. *Reversible Michaelis-Menten Gleichung mit zwei Inhibitor-Termen*

$$v = \frac{V_f \cdot \left(\frac{C_1}{K_{m,1}} \right) - V_r \cdot \left(\frac{C_2}{K_{m,2}} \right)}{1 + \left(\frac{I_1}{K_{i,1}} \right) + \left(\frac{C_1}{K_{m,1}} \right) + \left(\frac{C_2}{K_{m,2}} \right) \cdot \left(1 + \frac{I_1}{K_{i,1}} \right)} \cdot \left(\frac{1}{K_{i,2} + I_2} \right)^n \quad (\text{G-14})$$

a) $I_1 = \text{Citrat}$, $I_2 = 2\text{PG}$;15. *Reversible inhibierte Michaelis-Menten Gleichung*

$$v = \frac{V_f \cdot \left(\frac{C_1}{K_{m,1}} \right) - V_r \cdot \left(\frac{C_2}{K_{m,2}} \right)}{1 + \left(\frac{I}{K_i} \right) + \left(\frac{C_1}{K_{m,1}} \right) + \left(\frac{C_2}{K_{m,2}} \right) \cdot \left(1 + \frac{I}{K_i} \right)} \quad (\text{G-15})$$

a) $I = \text{AMP}$, b) $I = 2\text{PG}$, c) $I = \text{ATP}$, d) $I = \text{F6P}$;16. *Reversibles Geschwindigkeitsgesetz erster Ordnung*

$$v = k_1 \cdot C_1 - k_2 \cdot C_2 \quad (\text{G-16})$$

17. *Vereinfachte reversible Michaelis-Menten Gleichung*

$$v = \frac{k_1 \cdot C_1 - k_2 \cdot C_2}{k_3 + C_1/k_4 + C_2/k_5} \quad (\text{G-17})$$

18. *Vereinfachte reversible Michaelis-Menten Gleichung mit Inhibition*

$$v = \frac{k_1 \cdot C_1 - k_2 \cdot C_2}{(k_3 + C_1/k_4 + C_2/k_5) \cdot I}, \quad \text{a) } I = 2\text{PG} \quad (\text{G-18})$$

19. *Vereinfachte reversible Michaelis-Menten Gleichung mit exponentiellem Inhibitions-Term*

$$v = \frac{k_1 \cdot C_1 - k_2 \cdot C_2}{(k_3 + C_1/k_4 + C_2/k_5) \cdot \exp(k_6 \cdot I)}, \quad \text{a) } I = \text{Citrat} \quad (\text{G-19})$$

Tab. G–3 Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT120. *Irreversible Hill-Gleichung*

$$v = \frac{V_{\max} C^n}{K_s^n + C^n} \quad (\text{G-20})$$

21. *Irreversible Hill-Gleichung erweitert mit einem multiplikativen Aktivierungs-Term*

$$v = \frac{V_{\max} C^{n_1}}{K_s^{n_1} + C^{n_1}} \cdot \left(\frac{A}{K_a} \right)^{n_2} \quad (\text{G-21})$$

- a) $A = \text{F6P}$, b) $A = \text{PEP}$, c) $A = \text{ATP}$, d) $A = \text{G3P}$, e) $A = \text{Citrat}$, f) $A = \text{AMP}$
g) $A = \text{ADP}$

22. *Irreversible Hill-Gleichung für zwei Substrate*

$$v = \frac{V_{\max} \cdot C_1^{n_1}}{K_{s,1}^{n_1} + C_1^{n_1}} \cdot \frac{C_2^{n_2}}{K_{s,2}^{n_2} + C_2^{n_2}} \quad (\text{G-22})$$

23. *Irreversible Hill-Gleichung für zwei Substrate erweitert mit einem multiplikativen Inhibitions-Term*

$$v = \frac{V_{\max} \cdot C_1^{n_1}}{K_{s,1}^{n_1} + C_1^{n_1}} \cdot \frac{C_2^{n_2}}{K_{s,2}^{n_2} + C_2^{n_2}} \cdot \left(\frac{I}{K_i} \right)^{n_3}, \quad \text{a) } I = \text{PEP} \quad (\text{G-23})$$

24. *Irreversible Hill-Gleichung für drei Substrate erweitert mit einem multiplikativen Aktivierungs-Term*

$$v = \frac{V_{\max} \cdot C_1^{n_1}}{K_{s,1}^{n_1} + C_1^{n_1}} \cdot \frac{C_2^{n_2}}{K_{s,2}^{n_2} + C_2^{n_2}} \cdot \frac{C_3^{n_3}}{K_{s,3}^{n_3} + C_3^{n_3}} \cdot \left(\frac{A}{K_a} \right)^{n_4}, \quad \text{a) } A = \text{PEP} \quad (\text{G-24})$$

25. *Fluß in die Biosynthese*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{(K_{m,1} + C_1) \cdot (K_{m,2} + C_2)} \quad (\text{G-25})$$

26. *Irreversible Ordered Bi-Uni-Kinetik bzw. irreversible ordered Bi-Bi-Kinetik*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{K_{ia} \cdot K_{m,2} + K_{m,2} \cdot C_1 + C_1 \cdot C_2} \quad (\text{G-26})$$

27. *Reversible Ordered Uni-Bi-Kinetik*

$$v = \frac{V_f \cdot \left(C_1 - \frac{C_2 \cdot C_3}{K_{eq}} \right)}{K_{m,1} + C_1 \cdot \left(1 + \frac{C_2}{K_{i,2}} \right) + \frac{V_f}{V_r \cdot K_{eq}} \cdot (K_{m,3} \cdot C_2 + K_{m,3} \cdot C_3 + C_2 \cdot C_3)} \quad (\text{G-27})$$

28. *Reversible Ordered Uni-Bi-Kinetik erweitert mit einem multiplikativen Aktivierungs-Term*

$$v = \frac{V_f \cdot \left(C_1 - \frac{C_2 \cdot C_3}{K_{eq}} \right)}{K_{m,1} + C_1 \cdot \left(1 + \frac{C_2}{K_{i,2}} \right) + \frac{V_f \left(1 + \frac{A}{K_a} \right)^n}{V_r K_{eq}} (K_{m,3} C_2 + K_{m,3} C_3 + C_2 C_3)} \cdot \left(1 + \frac{A}{K_a} \right)^n \quad (\text{G-28})$$

- a) $A = \text{F6P}$

29. *Reversible Ordered Uni-Bi-Kinetik erweitert mit zwei multiplikativen Inhibitions-Termen*

$$v = \frac{V_f \cdot \left(C_1 - \frac{C_2 \cdot C_3}{K_{eq}} \right)}{K_{m,1} + C_1 \cdot \left(1 + \frac{C_2}{K_{i,2}} \right) + \frac{V_f}{V_r K_{eq}} (K_{m,3} C_2 + K_{m,3} C_3 + C_2 C_3)} \cdot \left(\frac{K_{i,1}}{I_1} \right)^{n_1} \cdot \left(\frac{K_{i,3}}{I_2} \right)^{n_2} \quad (\text{G-29})$$

- a) $I_1 = \text{GAP}$, $I_2 = \text{DHAP}$

Tab. G–3 Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT130. *Reversible Ordered Uni-Bi-Kinetik erweitert mit drei multiplikativen Aktivierungs-Termen*

$$v = \frac{V_f \cdot \left(C_1 - \frac{C_2 \cdot C_3}{K_{eq}}\right)}{K_{m,1} + C_1 \cdot \left(1 + \frac{C_2}{K_{i,2}}\right) + \frac{V_f}{V_r K_{eq}} (K_{m,3} C_2 + K_{m,3} C_3 + C_2 C_3)} \cdot \left(\frac{A_1}{K_{a1}}\right)^{n_1} \cdot \left(\frac{A_2}{K_{a2}}\right)^{n_2} \cdot \left(\frac{A_3}{K_{a3}}\right)^{n_3} \quad (G-30)$$

a) $A_1 = \text{PEP}, \quad A_2 = \text{Citrat}, \quad A_3 = \text{G3P}$

31. *Reversible Ordered Uni-Bi-Kinetik erweitert mit vier multiplikativen Inhibierungs- und zwei Aktivierungs-Termen*

$$v = \frac{V_f \cdot \left(C_1 - \frac{C_2 \cdot C_3}{K_{eq}}\right)}{K_{m,1} + C_1 \cdot \left(1 + \frac{C_2}{K_{i,2}}\right) + \frac{V_f}{V_r K_{eq}} (K_{m,3} C_2 + K_{m,3} C_3 + C_2 C_3)} \cdot \left(\frac{A_1}{K_{a1}}\right)^{n_1} \cdot \left(\frac{A_2}{K_{a2}}\right)^{n_2} \cdot \left(\frac{K_{i1}}{I_1}\right)^{n_3} \cdot \left(\frac{K_{i3}}{I_2}\right)^{n_4} \cdot \left(\frac{K_{i4}}{I_3}\right)^{n_5} \cdot \left(\frac{K_{i5}}{I_4}\right)^{n_6} \quad (G-31)$$

a) $A_1 = \text{PEP}, \quad A_2 = \text{Citrat}, \quad I_1 = \text{GAP}, \quad I_2 = \text{DHAP}, \quad I_3 = \text{ATP}, \quad I_4 = \text{E4P}$

32. *Irreversible Ordered Bi-Bi-Kinetik, ein Substrat inhibiert*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{\left(1 + \frac{C_1}{K_i}\right) \cdot K_{i,1} \cdot K_{m,2} + K_{m,2} \cdot C_1 + K_{m,1} \cdot C_2 \cdot \left(1 + \frac{C_1}{K_i}\right) + C_1 \cdot C_2} \quad (G-32)$$

a) $C_1 = \text{G6P}$

33. *Irreversible Ordered Bi-Bi-Kinetik, ein Produkt inhibiert*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{\left(1 + \frac{C_3}{K_i}\right) \cdot K_{i,1} \cdot K_{m,2} + K_{m,2} \cdot C_1 + K_{m,1} \cdot C_2 \cdot \left(1 + \frac{C_3}{K_i}\right) + C_1 \cdot C_2} \quad (G-33)$$

a) $C_3 = \text{NADPH}$

34. *Reversible Ordered Bi-Bi-Kinetik erweitert um einen multiplikativen Inhibitions-Term*

$$v = \frac{V_f \cdot \left(C_1 \cdot C_2 - \frac{C_3 \cdot C_4}{K_{eq}}\right)}{C_1 \cdot C_2 \cdot \left(1 + \frac{C_3}{K_{i,3}}\right) + K_{m,2} \cdot C_1 + K_{m,1} \cdot C_2 \cdot \left(1 + \frac{C_3}{K_{i,3}}\right) + \frac{C_2 C_3 C_4}{K_{i,2}} + A} \cdot \frac{1}{(K_i + I)^n} \quad (G-34)$$

$$A = K_{i,1} \cdot K_{m,2} + \frac{V_f}{V_r \cdot K_{eq}} \cdot \frac{1}{(K_i + I)^n} \cdot \left(K_{m,4} \cdot C_3 \cdot \left(1 + \frac{C_1}{K_i}\right) + C_4 \cdot K_{m,3} + C_3 \cdot C_4\right)$$

a) $I = \text{Gluc}, \quad \text{b) } I = \text{G6P}$

35. *Reversible Ordered Bi-Uni-Kinetik*

$$v = \frac{V_f (C_1 \cdot C_2 - \frac{C_3}{K_{eq}})}{C_1 \cdot C_2 + K_{m,1} \cdot C_2 + K_{m,2} \cdot C_1 + \frac{V_f}{V_r K_{eq}} \left(K_{m,3} + C_3 \left(1 + \frac{C_1}{K_{i,1}}\right)\right)} \quad (G-35)$$

36. *Reversible inhibierte Ordered Bi-Uni-Kinetik*

$$v = \frac{V_f (C_1 \cdot C_2 - \frac{C_3}{K_{eq}})}{C_1 \cdot C_2 + K_{m,1} \cdot C_2 + K_{m,2} \cdot C_1 + \frac{V_f}{V_r K_{eq}} \left(K_{m,3} + C_3 \left(1 + \frac{C_1}{K_{i,1}}\right)\right) \left(\frac{K_{i,1}}{I_1}\right)^{n_1} \cdot \left(\frac{K_{i,2}}{I_2}\right)^{n_2}} \quad (G-36)$$

a) $I_1 = \text{C5P}, \quad I_2 = \text{Citrat}; \quad \text{b) } I_1 = \text{G6P}, \quad I_2 = \text{F6P}$

Tab. G–3 Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT137. *Reversible aktivierte Ordered Bi-Uni-Kinetik*

$$v = \frac{V_f \left(C_1 \cdot C_2 - \frac{C_3}{K_{eq}} \right)}{C_1 \cdot C_2 + K_{m,1} \cdot C_2 + K_{m,2} \cdot C_1 + \frac{V_f}{V_r \cdot K_{eq}} \left(K_{m,3} + C_3 \left(1 + \frac{C_1}{K_{i,1}} \right) \right) \left(\frac{A}{K_a} \right)^n} \cdot \left(\frac{A}{K_a} \right)^n \quad (\text{G-37})$$

a) $A = \text{NADP}$; b) $A = \text{Pyr}$;38. *Reversible Ordered Bi-Bi-Kinetik*

$$v = \frac{V_f \cdot \left(C_1 \cdot C_2 - \frac{C_3 \cdot C_4}{K_{eq}} \right)}{C_1 \cdot C_2 \cdot \left(1 + \frac{C_3}{K_{i,3}} \right) + K_{m,2} \cdot C_1 + K_{m,1} \cdot C_2 \cdot \left(1 + \frac{C_3}{K_{i,3}} \right) + \frac{C_2 C_3 C_4}{K_{i,2}} + A} \quad (\text{G-38})$$

$$A = K_{i,1} \cdot K_{m,2} + \frac{V_f}{V_r \cdot K_{eq}} \left(K_{m,4} \cdot C_3 \cdot \left(1 + \frac{C_1}{K_i} \right) + C_4 \cdot K_{m,3} + C_3 \cdot C_4 \right)$$

39. *Reversible Ordered Bi-Bi-Kinetik erweitert um zwei multiplikative Inhibitions-Terme*

$$v = \frac{V_f \cdot \left(C_1 \cdot C_2 - \frac{C_3 \cdot C_4}{K_{eq}} \right)}{C_1 \cdot C_2 \cdot \left(1 + \frac{C_3}{K_{i,3}} \right) + K_{m,2} \cdot C_1 + K_{m,1} \cdot C_2 \cdot \left(1 + \frac{C_3}{K_{i,3}} \right) + \frac{C_2 C_3 C_4}{K_{i,2}} + A} \cdot \left(\frac{K_{i,5}}{I_1} \right)^{n_1} \cdot \left(\frac{K_{i,6}}{I_2} \right)^{n_1} \quad (\text{G-39})$$

$$A = K_{i,1} \cdot K_{m,2} + \frac{V_f}{V_r \cdot K_{eq}} \left(K_{m,4} \cdot C_3 \cdot \left(1 + \frac{C_1}{K_i} \right) + C_4 \cdot K_{m,3} + C_3 \cdot C_4 \right)$$

a) $I_1 = \text{G6P}$, $I_2 = \text{C5P}$; b) $I_1 = \text{GAP}$, $I_2 = \text{E4P}$ 40. *Irreversible Theorell-Chance-Bi-Bi-Gleichung*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{(K_{m,2} \cdot C_1 + K_{m,2} \cdot K_{i,1}) \cdot \left(1 + \frac{K-C_2}{K_i} \right) + C_2 \cdot C_1 + K_{m,1} \cdot C_2} \quad (\text{G-40})$$

41. *Irreversible Theorell-Chance-Bi-Bi-Gleichung erweitert mit einem multiplikativen Inhibitions-Term*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2}{(K_{m,2} \cdot C_1 + K_{m,2} \cdot K_{i,1}) \cdot \left(1 + \frac{K-C_2}{K_i} \right) + C_2 \cdot C_1 + K_{m,1} \cdot C_2} \cdot \left(\frac{I}{K_i} \right) \quad (\text{G-41})$$

a) $I = \text{NADP}$ 42. *Reversible Ping-Pong-Bi-Bi-Kinetik*

$$v = \frac{V_f \left(C_1 \cdot C_2 - \frac{C_3 C_4}{K_{eq}} \right)}{C_1 \cdot C_2 + K_{m,2} \cdot C_1 + K_{m,1} \cdot C_2 \cdot \left(1 + \frac{C_4}{K_{m,4}} \right) + \frac{V_f}{V_r \cdot K_{eq}} \left(K_{m,4} \cdot C_3 \cdot \left(1 + \frac{C_1}{K_{i,1}} \right) \right) + A} \quad (\text{G-42})$$

$$A = \frac{V_f \left(C_1 \cdot C_2 - \frac{C_3 \cdot C_4}{K_{eq}} \right)}{C_4 \cdot (K_{m,3} + C_3)}$$

43. *Irreversible Ter-Bi-Kinetik*

$$v = \frac{V_{\max} \cdot C_1 \cdot C_2 \cdot C_3}{K_{i,1} K_{i,2} K_{m,3} + K_{i,2} K_{m,3} C_3 + K_{i,1} K_{m,3} C_1 + K_{i,1} K_{m,2} C_2 + K_{m,2} C_2 C_3 + A} \quad (\text{G-43})$$

$$A = K_{m,1} C_1 C_2 + K_{m,3} C_1 C_3 + C_1 C_2 C_3$$

Tab. G–3 Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1

44. Goldbeter-Gleichung multiplikativ erweitert mit einer Michaelis-Menten-Kinetik

$$v = \frac{2 \cdot b \cdot D \cdot \frac{e}{e+1} \cdot C_1 \cdot (1 + C_3)^2 \cdot \left(1 + \frac{C_1}{e+1}\right)}{L \cdot (1 + C_1 \cdot c)^2 \cdot \left(1 + \frac{C_1}{e+1}\right)^2} \cdot \frac{C_2}{K_{m,2} + C_2} \quad (\text{G-44})$$

45. Goldbeter-Gleichung multiplikativ erweitert mit einer Michaelis-Menten-Kinetik und zwei Inhibitions-Termen

$$v = \frac{2 \cdot b \cdot D \cdot \frac{e}{e+1} \cdot C_1 \cdot (1 + C_3)^2 \cdot \left(1 + \frac{C_1}{e+1}\right)}{L \cdot (1 + C_1 \cdot c)^2 \cdot \left(1 + \frac{C_1}{e+1}\right)^2} \cdot \frac{C_2}{K_{m,2} + C_2} \cdot \left(\frac{K_{i,1}}{I_1}\right)^{n_1} \cdot \left(\frac{K_{i,2}}{I_2}\right)^{n_2} \quad (\text{G-45})$$

a) $I_1 = \text{PEP}$, $I_2 = \text{ADP}$, b) $I_1 = \text{ADP}$, $I_2 = \text{FBP}$

46. Kinetik für die Phosphofruktokinase

$$v = \frac{K_{10}}{1 + \exp(K_9 \cdot C_1)} \cdot \frac{K_7 C_2 C_3 - K_8 C_1 C_4}{1 + K_1 C_2 + K_2 C_3 + K_3 C_3 C_2 + K_4 C_4 + K_5 C_1 + K_6 C_4 C_2} \quad (\text{G-46})$$

47. Kinetik für die Phosphofruktokinase nach (Theobald et al. 1997a)

$$v = \frac{V_{\max} C_1}{C_1 + K_1 \cdot \left(1 + \frac{C_2}{K_2}\right)} \cdot \frac{C_3}{C_3 + K_9 \cdot \frac{1 + C_1/K_4 + C_2/K_5 + M/K_6}{1 + C_2/K_7 + M/K_8}} \cdot A \quad (\text{G-47})$$

$$A = \frac{1}{1 + \left(1 + \frac{I}{K_i}\right)^4} \cdot \frac{K_3}{1 + \frac{C_3}{K_9 \cdot \frac{1 + C_1/K_4 + C_2/K_5 + M/K_6}{1 + C_2/K_7 + M/K_8}}} \quad (\text{G-47})$$

48. Allosterische Inhibition

$$v = \frac{V \cdot \left(1 + \frac{C_1}{K_{m,1}}\right)^{n-1}}{L \cdot \left(1 + \frac{I}{K_i}\right)^n + \left(1 + \frac{C_1}{K_{m,1}}\right)^n} \quad (\text{G-48})$$

a) $I = \text{PEP}$, b) $I = \text{ADP}$, c) $I = \text{NADP}$

49. Nicht-exklusive Substrat- und Effektor-Bindung

$$v = \frac{V_{\max} \cdot c \cdot \frac{C_1}{K_{m,1}} \cdot \left(1 + c \frac{C_1}{K_{m,1}}\right)^{n-1} \cdot \left(1 + e \frac{I}{K_i}\right)^n + \frac{C_1}{K_{m,1}} \cdot \left(1 + \frac{C_1}{K_{m,1}}\right)^{n-1} \cdot \left(1 + \frac{I}{K_i}\right)^n}{L \cdot \left(1 + c \frac{C_1}{K_{m,1}}\right)^n \cdot \left(1 + e \frac{I}{K_i}\right)^n + \left(1 + \frac{C_1}{K_{m,1}}\right)^n \cdot \left(1 + \frac{I}{K_i}\right)^n} \quad (\text{G-49})$$

a) $I = \text{PEP}$

50. Exklusive Substrat-Bindung, K-System, 2 Inhibitoren

$$v = \frac{V_{\max} \left(\frac{C_1}{K_{m,1}} \cdot \left(1 + \frac{C_1}{K_{m,1}}\right)^{n-1}\right)}{L \cdot \left(1 + \frac{I_1}{K_{i,1}} + \frac{I_2}{K_{i,2}}\right)^n + \left(1 + \frac{C_1}{K_{m,1}}\right)^n} \quad (\text{G-50})$$

a) $I_1 = \text{PEP}$, $I_2 = \text{FBP}$; b) $I_1 = \text{PEP}$, $I_2 = \text{AMP}$; c) $I_1 = \text{Citrat}$, $I_2 = \text{PEP}$

51. Geschwindigkeitsgesetz erster Ordnung für eine bimolekulare reversible Reaktion

$$v = k \cdot C_1 \cdot C_2 \quad (\text{G-51})$$

52. Geschwindigkeitsgesetz n-ter Ordnung für eine bimolekulare reversible Reaktion

$$v = k_f \cdot C_1^{n_1} \cdot C_2^{n_2} - k_r \cdot C_3^{n_3} \cdot C_4^{n_4} \quad (\text{G-52})$$

Tab. G–3 Reaktionskinetische Terme für die Modelle zur Beschreibung des Experimentes WT1

53. *Geschwindigkeitsgesetz n-ter Ordnung für die Hinreaktion, Hill-Gleichung für die Rückreaktion*

$$v = k_1 \cdot C_1^{n_1} - \frac{V_{\max} \cdot C_2^{n_2}}{K_s^{n_2} + C_2^{n_2}} \tag{G-53}$$

54. *Reversible Ordered Uni-Bi erweitert mit einem multiplikativen Inhibitions-Term*

$$v = \frac{V_f \left(C_1 - \frac{C_2 \cdot C_3}{K_{eq}} \right)}{K_{m,1} + C_1 \cdot \left(1 + \frac{C_2}{K_{i,2}} \right) + \frac{V_f}{V_r K_{eq}} \cdot (K_{m,3} \cdot C_2 + K_{m,2} \cdot C_3 + C_2 \cdot C_3)} \cdot \frac{1}{I^n + K_i^n} \tag{G-54}$$

a) $I = \text{F6P}$

55. *Goldbeter-Gleichung multiplikativ erweitert mit einer Michaelis-Menten-Kinetik, zwei Aktivierung- und einem Inhibitions-Term*

$$v = \frac{2 \cdot b \cdot D \cdot \frac{e}{e+1} \cdot C_1 \cdot (1 + C_3)^2 \cdot \left(1 + \frac{C_1}{e+1} \right)}{L \cdot (1 + C_1 \cdot c)^2 \cdot \left(1 + \frac{C_1}{e+1} \right)^2} \cdot \frac{C_2}{K_{m,2} + C_2} \cdot \left(\frac{K_{i,1}}{I_1} \right)^{n_1} \cdot \left(\frac{K_{a,1}}{A_1} \right)^{n_2} \cdot \left(\frac{K_{a,2}}{A_2} \right)^{n_3} \tag{G-55}$$

a) $I_1 = \text{PEP}, \quad A_1 = \text{ADP}, \quad A_2 = \text{F6P}$

Tab. G–4 Modelle der Topologien 0.a (links) und 1 (rechts); für die Reaktionsschritte verwendete Kinetiken

		Reaktions-Nr.	Modell-Nr. (.) Kinetik-Nr.
Reaktions-Nr.	Modell-Nr. (.) Kinetik-Nr	I.1	1-4.7
0.a.1	72-73.7	I.2	1-4.2
0.a.2	71-72.11;	I.3	1.47; 2-4.15a
0.a.3	71-72.38	I.4	1-4.20
0.a.4	71-73.2	I.5	1.11; 2-4.2
0.a.5		I.6	1-4.20
0.a.6	71-73.40	I.7	1-4.10
0.a.7		I.8	1.32a; 2.4a; 3.33a; 4.2
		I.9	1.25
		I.10	1-4.27
		I.11	1-4.43

Tab. G-5 Für die verschiedenen Reaktionsschritte verwendete Kinetiken in den Modellen der Topologie 2

Reaktions-Nr.	Modell-Nr. (.) Kinetik-Nr.
II.1	5-16, 19-25.7; 17-18.8a
II.2	5-17, 19-25.2; 18.11
II.3	5.50a; 6, 19, 20.15a; 7,14,21.48a; 8.50b; 9-12,22,23.49a; 13.47; 15-18.45a; 24.48b; 25.23a
II.4	5-7,9,10,18-20,22,23,25.20; 8,24.15b; 13-17.2; 11,12.11
II.5	5-10,16,19-25.2; 11,12.20; 13.11; 14.6a; 15,17,18.13a
II.6	16-18.11
II.7	5-25.20
II.8	5-8,12-21,24,25.10; 9,10,22,23.2; 11.20
II.9	14.11
II.10	10-25.20
II.11	10,11,23-25.38, 12-18.15.c
II.12	10,11.7; 12-18.11; 23-24.43; 25.11
II.13	10,24.25; 11-18,23,25.20
II.14	5-8,19,21.5a; 9-13,15-18,20,22-25.40; 14.41a
II.15	5-25.11
II.16	13-18.20
II.17	5-25.27
II.18	5-25.38
II.19	5-14,19-25.43, 15-18.7

Tab. G-6 Für die verschiedenen Reaktionsschritte verwendete Kinetiken in den Modellen der Topologie 3

Reaktions-Nr.	Modell-Nr. (.) Kinetik-Nr.
III.1	26-29,43,47,51,58.7; 30,52.8b; 31-42,44-46,48-50,53-57,74,76.8a
III.2	26-29,47.2; 30,31,40-44,48,49,52,58,74.11; 32-34,36-39,45,46,50,51.15d; 35.5a; 53,55-57,76.15d; 54.20
III.3	26.50a; 27.48a; 28.50c; 29.49a; 32-45,47-50,53,55-57.45a; 30-31,52,54,58,74.44; 51.11; 46.20; 76.45b
III.4	26,27,30-39,43,45-51,52-58,74,76.20; 28,29.11; 40-42,44.2
III.5	48,49.20
III.6	26.2; 28,29,33,40,47-49,54,58.27; 27.2; 30.30a; 31,32,52,74.31a; 34-39, 41-46,50,51,55,56.29a; 53,76.29a; 57.1
III.7	26-33,37-51,74.11; 34,76.18a; 35.19a; 36.14a; 52-56, 58.11, 57.1;
III.8	30.15c; 40,46, 47.20,58.20; 49.11; 54.20; 57.21c
III.9	30,46,49,57.20
III.10	26-33,36-39,44,47-58,74.20; 34,35,76.1; 40-43,45,46.2
III.11	26,27,30,33-57,76.10; 32,74.7a; 28-29.20; 31.7a;58.2
III.12	30-37,40,47-51, 53-58,76.38; 38,39.39a; 41-46.34b; 52,74.12
III.13	30,41,42,45.11
III.14	30.11; 31-46,48-56,74,76.35; 47.7c; 57,58.27
III.15	30-47,76.20; 53-58.20
III.16	26-28.5a; 29-45,48-51,53,55-57,76.40; 46.48c; 54.8c; 47,52,74.8a
III.17	26-39,47-51,53-58,76.11; 40.36a; 41-44.37a; 45.7; 46,58.35; 52,74.26; 58.35
III.18	49,57.21
III.19	30-31, 39,40-45.20; 32-38,46,50,51,53,76.21a
III.20	39.1; 40-45.7b; 46.21a; 47-49,54,56,57,58.20; 55.21a
III.21	26-42,44,46-58,76.27; 43.28a; 45.20; 74.21f
III.22	26-39,46-51,53-58,76.38; 45.26; 40-44.39b; 52,74.9a ;
III.23	26-37,39,52,53,55.43; 54,74,76.7; 47-51.10, 57,58.7;38.24a
III.24	41.34a

Tab. G-7 Für die verschiedenen Reaktionsschritte verwendete Kinetiken in den Modellen der Topologien 4 (links) und 5 (rechts)

Reaktions-Nr.	Modell-Nr. (.) Kinetik-Nr.	Reaktions-Nr.	Modell-Nr. (.) Kinetik-Nr.
IV.1	59-60.8a; 61.21b; 62.7	V.1	63.7; 64-68.8a; 69.51; 70,75.52
IV.2	59.11; 60-62.11	V.2	63.11; 64-68.15d; 69.11; 70.53; 75.51
IV.3	59,60.45a; 61,62.44	V.3	67,68.20 ; 69.7c; 70,75.7c
IV.4	59-62.20	V.4	63-68.45a ;69.45a; 70.45a
IV.5	59-62.27	V.5	63-68.20 ;69.20; 70.15b ; 75.55a
IV.6	59-62.11	V.6	63.27; 64-66.29a; 67.1; 68.29a ; 69,70,75.21d
IV.7	61,62.20	V.7	63-66,68.11; 67.1 ; 69,70,75.11
IV.8	59-62.20	V.8	67,68.11 ; 69,70,75.21e
IV.9	59-61.10, 62.2	V.9	67,68.20 ;69,70.21d ; 75.21a
IV.10	59-62.38	V.10	63-68.10 ; 69,70,75.10
IV.11	59-62.35	V.11	63-68.20 ; 69,70,75.20
IV.12	59-62.20	V.12	63-68.38;69,75.39a; 70.38
IV.13	59.40; 60-62.8a	V.13	63-68.20;
IV.14	59.62.20	V.14	63-68.35; 69.35; 70.37b; 75.8d
IV.15	59-62.35	V.15	63-70,75.40
IV.16	59-62.20	V.16	63-68.11; 69,70,75.36b
IV.17	59-62.27	V.17	63-68.27; 69,75.27; 70.54a
IV.18	59-62.42	V.18	63.20 ; 64-68.21a; 69.21c; 70.21c; 75.21g
IV.19	59-62.38	V.19	65.38;66-68,75.7;
IV.20	59-62.7	V.20	63,64.7; 69.7
		V.21	69.8a; 70.8a ; 75.7
		V.22	69.7 ; 75.20
		V.23	69,70,75.26;
		V.24	69,70,75.21a

Tab. G-8 Zuordnung der Modellnummern aus Tabellen G-4 bis G-7 zu den Namen bzw. der Nummern der Modelle im Hauptteil

Nr. im Hauptteil	Nr. im Anhang	Name: Modellierungsgruppe/Modell
1	34	E.coli6/VersionV
2	36	E.coli6/VersionVII
3	37	E.coli6/VersionVIII
4	53	E.coli6/VersionIII
5	55	E.coli6/VersionIIIa
6	56	E.coli6/VersionIIIc
7	74	modgroup89/model45
8	65	E.coli6/VersionXXI
9	66	E.coli6/VersionXXII
10	69	pts(+) K12 Gluc/Gluc/ID 101 + erweiterte PEP-Rkt verbessert
11	70	Weiterführung von ID 26/ID 131 veränderte PFK-2 Reaktion
12	75	Weiterführung von ID 26/ID 101 + erweiterte PEP Reaktion

Forschungszentrum Jülich
in der Helmholtz-Gemeinschaft



Jül-4147
September 2004
ISSN 0944-2952