

Tandem-Mass-Spectrometry-Driven Investigation of
the Anaplerotic Reactions in *Corynebacterium*
glutamicum

Tandem Massenspektrometrie getriebene
Untersuchung der anaplerotischen Reaktionen in
Corynebacterium glutamicum

Von der Fakultät für Maschinenwesen der Rheinisch-Westfälischen Technischen
Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der
Naturwissenschaften genehmigte Dissertation

vorgelegt von

Jannick Kappelmann

Berichter: Universitätsprofessor Dr. rer. nat. Wolfgang Wiechert
 Universitätsprofessor Dr.-Ing. Ralf Takors

Tag der mündlichen Prüfung: 23. Februar 2018

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Meinem liebsten Ivo Schatz

Statement Of Authorship

Hereby I confirm that the content of this thesis is my own work, and all sources and aids that were used to generate the content of this thesis are properly referenced. The content of this thesis has not been published by or submitted to any other institution.

.....
Place, Date

Zusammenfassung

Die anaplerotischen Reaktionen bilden ein zyklisches Reaktionsnetzwerk um die Metabolite Phosphoenolpyruvat, Pyruvat, Oxaloacetat und Malat, dessen Carboxylierungs- und Decarboxylierungsreaktionen eine Verbindung zwischen Glykolyse und Tricarbonsäure (TCA)-Zyklus herstellen. Die Netto-Aktivität der anaplerotischen Carboxylierung ist essentiell zur Bereitstellung von TCA- Zyklus Intermediaten und ist daher limitierend für die Biosynthese von aus letzteren abgeleiteten Aminosäuren, insbesondere L-Lysin und L-Glutamat. Der Großteil der Weltproduktion dieser beiden Aminosäuren wird durch die biotechnologische Nutzung von *Corynebacterium glutamicum* bestritten, einem Arbeitspferd der Biotechnologie und Modellorganismus der Systembiologie gleichermaßen.

Die herausragende Bedeutung der anaplerotischen Reaktionen für die biotechnologische Nutzung von *C. glutamicum* motiviert diese Arbeit. Einleitend klärt eine theoretische Untersuchung die Ursache für erfolglos bestimmte anaplerotische Aktivitäten mittels ^{13}C -Metabolischer Stoffflussanalyse (^{13}C -MFA) in vorherigen Studien auf. Unter Nutzung einer fokussierten Isotopomer-Bilanz des anaplerotischen Reaktionsnetzwerkes kann ohne Beschränkung der Allgemeinheit gezeigt werden, dass für *C. glutamicum* lediglich Deletionsmutanten in anaplerotischen Reaktionen eine eindeutige Bestimmung der anaplerotischen Flüsse aus stationären isotopischen ^{13}C -Markierungsdaten zulassen.

Dieser Befund erzwang die Untersuchung einer Sammlung von Einzel- und Doppeldeletionsmutanten, deren Metabolismus mittels Triple Quadrupol-Massenspektrometrie (QqQ MS) untersucht wurde. Mittels Quadrupole-Flugzeit-Massenspektrometrie (QqTOF MS) wurde ein Proteom-Datensatz erzeugt, aus dem unter Nutzung eines Regressionsansatzes präzise relative Änderungen für ein knappes Drittel der Proteine von *C. glutamicum* geschätzt wurden. Eine wesentliche biologische Erkenntnis aus der Gesamtheit der Daten besteht in dem rigorosen Nachweis, dass der Glyoxylat-Zyklus als anaplerotische Reaktion unter ausschließlich glykolytischen Bedingungen in *C. glutamicum* aktiv sein kann, was durch eine Metabolit-Protein Interaktion und keine Mutation der relevanten Regulatoren vermittelt wird.

Der Befund einer inkonsistenten Flusschätzung im Rahmen einer ^{13}C -MFA der untersuchten Stämme veranlasste die Entwicklung einer neuen QqTOF MS-basierten Akquisitionsmethode, die die Kollisions-induzierte Dissoziation (CID) von Anionen des Zentralstoffwechsels und Kationen von Aminosäuren zur positionsaufgelösten ^{13}C -Markierungsmessung nutzt. Zu diesem Zweck wurden akkurate Massenspektrometrie, selektiv-markierte Standards und publizierte Fragmentierungswege genutzt, um alle dominanten Massenpeaks einer großen Auswahl von Metaboliten Metabolitauswahl strukturell zu annotieren. Die vorgenommene Analyse stellt die detaillierteste Beschreibung des Verbleibs von Kohlenstoffatomen in LC-ESI-MS/MS Fragmenten dar. Über die Positionsauflösung hinaus, kann gezeigt werden, dass die CID von Aminosäuren dazu genutzt werden kann, ^{13}C - und ^{15}N -Massenisotopomere dieser Metabolitklasse analytisch zu trennen, ohne sie in der m/z -Domäne aufzulösen.

Abstract

The anaplerotic reactions form the basis of a circular reaction network around the metabolites phosphoenolpyruvate, pyruvate, oxaloacetate, and malate, comprising specific carboxylation and decarboxylation reactions which link glycolysis to the tricarboxylic acid cycle. The net carboxylation activity of anaplerotic reactions is essential for the replenishment of tricarboxylic acid cycle intermediates and therefore directly limits the biosynthetic flux towards amino acids derived from it, most notable L-lysine and L-glutamate. The microbial production of these amino acids is primarily accounted for by the biotechnological usage of *Corynebacterium glutamicum*, a major workhorse of industrial biotechnology and model organism of systems biology alike.

The pivotal role of anaplerotic reactions in the biotechnological usage of *C. glutamicum* motivates the present thesis which is dedicated to a thorough study of these reactions in this organism. First, a theoretical analysis establishes the reason for the unsuccessfully determined anaplerotic reaction activities by ^{13}C -Metabolic Flux Analysis (^{13}C -MFA) in previous studies. Using a focused isotopomer network of the anaplerotic node, it is shown in full generality that for *C. glutamicum* only certain anaplerotic deletion mutants allow to uniquely determine the anaplerotic fluxes from stationary ^{13}C -isotopomer data. This result prompted the investigation of a set of single and double deletion mutants whose metabolome was investigated by a protocol employing a liquid chromatography electrospray ionization triple quadrupole mass spectrometry (LC-ESI-QqQ MS) platform. By means of quadrupole-time-of-flight technology (QqTOF) a proteome data set for the investigated deletion mutants was generated and evaluated with a novel regression approach, yielding precise fold changes for almost a third of all open reading frames of *C. glutamicum*. A major biological finding arising from the total data set consists in the proof that the glyoxylate shunt can be active as anaplerotic reaction in *C. glutamicum* under solely glycolytic conditions, which is brought about by a metabolite-regulator interaction.

The finding of inconsistently estimated fluxes in a ^{13}C -Metabolic Flux Analysis of the investigated strains prompted the development of a novel LC-ESI-QqTOF MS-based ^{13}C -profiling method, which harvests the collision induced dissociation (CID) of product anions of central metabolism as well as product cations of amino acids for the positional resolution of ^{13}C -label enrichment. For this purpose, accurate mass spectrometry, selectively-labeled standards, and published fragmentation pathways are used to structurally annotate all dominant mass peaks of a large collection of metabolites, resulting in the most detailed map of the carbon atom fate of LC-ESI-MS/MS collisional fragments yet. Apart from providing positional resolution of ^{13}C -enrichment, it is shown that the CID of amino acids can be harvested to separate ^{13}C - and ^{15}N -isotopic species of these metabolites without resolution in the m/z domain.

Acknowledgements

The present thesis is the culmination of my research at the Institute for Bio- and Geosciences I of Forschungszentrum Jülich, carried out between January 2014 and March 2017. During this time, I greatly benefited from the help of other colleagues, to some of whom I would like to explicitly express my gratitude. First and foremost, I would like to thank the supervisor of my doctoral thesis, Prof. Dr. Wolfgang Wiechert. Without his guidance in the *Studienkolleg* of the German National Academic Foundation my interest in quantitative biology and systems biology would not have been stimulated in the first place. For his mentorship previous to and during my PhD work I owe him a great debt. Equally as important for my work was the collaboration with Dr. Stephan Noack. The countless fruitful discussions with him provided crucial guidance for my work. At all times, I could reflect my ideas with him, no matter how flawed they may have been. Thirdly, I would like to thank Prof. Dr. Ralf Takors for his commitment to assess my thesis as second reviewer. The constructive collaboration with Dr. Bastian Blombach made the experimental part of this work possible. Dr. Bastian Blombach constructed all the deletion strains investigated in this thesis. Furthermore, I owe a great debt to Prof. Dr. Marco Oldiges, chair of the Bioprocesses and Bioanalytics group, who made the acquisition of the QqTOF MS device used in this thesis possible. I would like to thank Bianca Klein for her technical assistance in the course of my work. Not only did Bianca prepare and measure the samples for the proteome analysis, she also helped with any technical issues of the HPLC systems and the QqTOF system. Bianca furthermore diligently carried out the direct infusion experiments on the QqTOF system. This work would not have been possible without her. Petra Geilenkirchen measured samples from the metabolome and ^{13}C -labeling experiments, for which I would like to thank her. I would also like to express my gratitude towards my students Eva Höllmüller and Mathias Papenfuß. Eva Höllmüller assisted in carrying out the Pyruvate Carboxylase assay and the L-Glutamic Dehydrogenase assay. Mathias Papenfuß implemented the data storage framework for proteome data in MATLAB.

Published Work

Publications

- Kappelmann, J.; Wiechert, W. and Noack, S. (2016) Cutting the Gordian Knot: Identifiability of Anaplerotic Reactions in *Corynebacterium glutamicum*. *Biotechnology and Bioengineering* 113(3): 661-674.
- Kappelmann, J.; Klein, B.; Geilenkirchen P. and Noack, S. (2017) Comprehensive and Accurate Tracking of Carbon Origin of LC-Tandem Mass Spectrometry Collisional Fragments for ^{13}C -MFA. *Analytical and Bioanalytical Chemistry* 409(9):2309-2326
- Kallscheuer, N.; Vogt, M.; Kappelmann, J.; Krumbach, K.; Noack, S.; Bott, M; Marienhagen, J. (2016) Identification of the *phd* gene cluster responsible for phenylpropanoid utilization in *Corynebacterium glutamicum*. *Applied Microbiology and Biotechnology* 100(4):1871-81.
- Hemmerich, J.; Tenhaef, N; Steffens, C.; Kappelmann, J. et al. (2018) Less Sacrifice, More Insight: Repeated Low-Volume Sampling of Microbioreactor Cultivations Enables Accelerated Deep Phenotyping of Microbial Strain Libraries. *Biotechnology Journal* article in press

Oral Presentations

- Kappelmann, J.; Noack, S. Interpreting LC-MS/MS spectra from primary metabolites for tandem MS-based ^{13}C -Fluxomics. 21st International Mass Spectrometry Conference (IMSC 2016), Toronto, Canada, 22nd-26th August 2016
- Noack, S.; Kappelmann, J. Interpretation von LC-MS/MS-Spektren von Primärmetaboliten zur Tandem-MS-basierten ^{13}C -Stoffflussanalyse. 7. LC-MS/MS Symposium, Berlin, Germany, 15th March 2017

Posters

- Wiechert, W.; Noack, S.; Kappelmann, J. ^{13}C -Metabolic Flux Analysis at its Limits: Identifiability of anaplerotic reactions in *Corynebacterium glutamicum*. Metabolic Engineering 11 Conference (ME11), Kobe, Japan, 26th-30th June 2016

In Preparation

- Kappelmann, J.; Noack, S. (2019) Isobaric Deconvolution of Reflectron TOF Mass Peaks. in preparation to be submitted as technical note to the *Journal of The American Society for Mass Spectrometry*
- Kappelmann, J.; Beyß, M.; Nöh, K.; Wiechert, W.; Noack, S. (2019) Separation of ^{13}C - and ^{15}N -Isotopologues of Amino Acids without Mass Resolution by means of O-Phthalaldehyde Derivatization and Collision Induced Dissociation. In preparation to be submitted as research article to *Analytical Chemistry*

- Kappelmann, J.; Klein, B.; Papenfuß, M.; Pooth, V.; Lange, J.; Blombach, B.; Takors, R.; Polen, T.; Oldiges, M.; Noack, S. (2019) A Linear Mixed Effects Model for SWATH-MS data accounting for peptide correlation. In preparation to be submitted as research article to *Journal of Proteome Research*
- Kappelmann, J.; Wiechert, W.; Noack, S. (2019) Anaplerotic reactions: pivotal in function, yet inconsistent in definition. In preparation to be submitted as Current Opinion to *FEMS Microbiology Letters*

Contents

Zusammenfassung	v
Abstract	vi
Acknowledgements	vii
Published Work	viii
Nomenclature	xiii
1 Introduction	1
1 Motivation	1
2 Outline of this thesis	3
2 Background	4
3 Anaplerotic Reactions and their Role in the Central Carbon Metabolism	4
4 LC-ESI-MS/MS instrumentation	13
4.1 Motivation	13
4.2 LC-ESI-QqQ MS	13
4.3 QqTOF MS technology	18
5 LC-ESI-MS/MS-based quantitative Biology	21
5.1 LC-ESI-MS/MS-based Metabolomics	21
5.2 LC-ESI-MS/MS-based Proteomics	22
5.3 ¹³ C-Fluxomics	27
3 Materials and Methods	44
6 Materials	44
6.1 Strain maintenance	44
6.2 Chemicals	44
6.3 Isotopic standards	44
6.4 Consumable solutions and chemicals	46
6.5 Media	47
6.6 Devices and Chromatographic Material	49
6.7 Software	51
7 Methods	51
7.1 Bioreactor cultivations	51
7.2 Metabolome analysis	52
7.3 BioLector cultivations	54
7.4 Analytics for extracellular phenotyping	54

7.5	LC-ESI-QqQ analytics for central carbon metabolism intermediates	56
7.6	SWATH acquisition method	57
7.7	SWATH data processing	58
7.8	Generation of product ion spectral libraries on a QqQ MS and QqTOF MS	59
7.9	<i>In vitro</i> enzyme assays to generate selectively labeled ¹³ C-species and ¹⁸ O-species	59
7.10	Estimation of extracellular rates	61
7.11	Full information FIA	62
4	Results and Discussion - Part I	64
8	Structural Identifiability Analysis of anaplerotic reactions by ¹³ C-MFA	64
8.1	Motivation	64
8.2	Network reduction for exact flux identifiability analysis	64
8.3	Full information FIA in a focused network model	67
8.4	Partial information FIA in a focused network model	71
8.5	Non-unique global FIA in a comprehensive isotopomer model of <i>C. glutamicum</i>	73
8.6	Conclusions	77
9	Extracellular phenotyping of anaplerotic deletion mutants	77
9.1	Motivation	77
9.2	Adaptation of the Δ(PCx PEPCx) strain to glucose as sole carbon source	78
9.3	Specific extracellular rates of anaplerotic deletion mutants	83
9.4	Bicarbonate pool as carbon sink	89
9.5	Summary	96
10	Metabolome analysis	97
10.1	Intracellular metabolome	97
10.2	Exometabolome	102
11	Proteome analysis	103
11.1	Data generation and data handling	103
11.2	Consistency and data quality check	104
11.3	Fold change estimation using Poisson regression	106
11.4	Fold changes of selected anaplerotic enzymes	114
11.5	Fold change estimation in anaplerotic deletion mutants	115
12	¹³ C-Metabolic Flux Analysis	121
12.1	¹³ C-Labeling experiment and enrichment profiling	121
12.2	Flux estimation	123
5	Results and Discussion - Part 2	129
13	QqTOF MS technology for ¹³ C-isotopomer profiling	129
13.1	Motivation	129
13.2	¹³ C-profiling method	129

13.3	Carbon identity elucidation of CID fragments of primary metabolites	132
14	QqTOF MS technology for simultaneous ^{13}C - and ^{15}N -isotope enrichment profiling . .	151
14.1	Motivation	151
14.2	Problem statement	151
14.3	Analytical separation of ^{13}C - and ^{15}N -MI by means of CID	152
14.4	Spectral deconvolution of reflectron TOF mass peaks for ^{13}C - and ^{15}N -MI separation	157
14.5	Validation using defined ^{13}C - and ^{15}N -MI mixtures	160
6	Conclusions And Outlook	163
	Appendices	172
1	Primers used to verify deletion mutant identity	172
2	Correlation between BV concentration and CDW concentration	172
3	Model-based estimates for specific rates in anaplerotic deletion mutants	173
4	Gaussian Error Propagation	174
5	Parameter values for the simulation study of CO_2 dynamics	175
6	Q1 Windows for SWATH Acquisition	176
7	Exemplary Design Matrix	177
8	Sum of Squared Residuals from the flux estimation in a <i>C. glutamicum</i> mutant library .	179
9	HPLC method for a QqTOF acquisition method	180
10	Exemplary acquisition of an amino acid standard mixture on the QqTOF device	181
	Bibliography	189

Nomenclature

	Abbreviations
^{13}C -MFA	^{13}C -Metabolic Flux Analysis
ANOVA	analysis of variance
CAD	collision gas pressure parameter
CE	collision energy
CI	chemical ionization
CID	collision-induced dissociation
CUR	curtain gas flow parameter
CXP	cell exit potential
DP	declustering potential
EI	electron ionization
EMP pathway	Embden-Meyerhof-Parnas pathway
ESI	electrospray ionization
FA	fractional abundance
FC	fold change
FDR	false discovery rate
FIA	flux identifiability analysis
FTIR	Fourier transform interference spectroscopy
FWHM	full width at half maximum
GC	gas chromatography
GS1	ion source gas I flow parameter
GS2	ion source gas II flow parameter
HPLC	high performance liquid chromatography
IQR	interquartile range
IS	internal standard
L/H ratio	light-to-heavy ratio
LC	liquid chromatography
MI	mass isotopomer
MID	mass isotopomer distribution
MRM	multiple reaction monitoring
MS/MS	tandem mass spectrometry
MTP	microtiter plate
ORF	open reading frame
PCA	principle component analysis
PP pathway	pentose phosphate pathway
Q1	first quadrupole

Q2	second quadrupole (collision cell)
Q3	third quadrupole
QqQ	triple quadrupole
QqTOF	quadrupole TOF
rf	radiofrequency
RP _{FWHM}	resolving power (full width at half maximum)
SNP	single nucleotide polymorphism
SRM	selected reaction monitoring
SSR	sum of squared residuals
SVD	singular value decomposition
SWATH	Sequential Window Acquisition of all THeoretical fragment-ion spectra
TBDMS	<i>tert</i> -butyldimethylsilyl
TCA cycle	tricarboxylic acid cycle
TEM	ion source temperature parameter
TMI	tandem mass isotopomer
TMID	tandem mass isotopomer distribution
TMS	trimethylsilyl
TOF	time of flight

Metabolites	
α KG	α -ketoglutarate
6PG	6-phosphogluconate
ACCOA	acetyl coenzyme A
ACO	aconitate
ALA	L-alanine
ARG	L-arginine
ASN	L-asparagine
ASP	L-aspartate
CIT	citrate
COA	coenzyme A
F6P	fructose 6-phosphate
FADH ₂	reduced form of flavin adenine dinucleotide
FBP	fructose 1,6-bisphosphate
G6P	glucose 6-phosphate
GLN	L-glutamine
GLU	L-glutamate
GLX	glyoxylate
GLY	L-glycine
GTP	guanosine 5-triphosphate

HOM	L-homoserine
ICIT	isocitrate
IMP	inosine 5-monophosphate
ISOLEU	L-isoleucine
LAC	L-lactate
LEU	L-leucine
LYS	L-lysine
MAL	L-malate
MET	L-methionine
MQ ₂	menaquinol
NAD ⁺	oxidized form of nicotinamide adenine dinucleotide
NADH	reduced form of nicotinamide adenine dinucleotide
NADP ⁺	oxidized form of nicotinamide adenine dinucleotide phosphate
NADPH	reduced form of nicotinamide adenine dinucleotide phosphate
OAA	oxaloacetate
OAS	O-acetyl-L-serine
ORN	L-ornithine
PEP	phosphoenolpyruvate
PGA	3-phosphoglycerate
PHE	L-phenylalanine
PRO	L-proline
PYR	pyruvate
S7P	sedoheptulose 7-phosphate
SCOA	succinyl coenzyme A
SER	L-serine
SUC	succinate
THR	L-threonine
TYR	L-tyrosine
UMP	uridine 5-monophosphate
VAL	L-valine

Proteins

6PGL	6-Phosphogluconolactonase
ACK	Acetate kinase
ACN	Aconitate hydratase
ASAT	Aspartate transaminase
CS	Citrate synthase
ENO	Enolase
FBA	Fructose-bisphosphate aldolase

FBPase	Fructose 1,6-bisphosphatase
Fum	Fumarase
G6PDH	Glucose 6-phosphate dehydrogenase
GAPDH	Glyceraldehyde phosphate dehydrogenase
ICD	Isocitrate dehydrogenase
ICL	Isocitrate lyase
LDH	Lactate dehydrogenase
MDH	Malate dehydrogenase
ME	Malic enzyme
MQO	Malate:quinone oxidoreductase
MS	Malate synthase
ODx	Oxalacetate decarboxylase
PCx	Pyruvate carboxylase
PDHC	Pyruvate dehydrogenase complex
PEPCK	Phosphoenolpyruvate carboxykinase
PEPCx	Phosphoenolpyruvate carboxylase
PFK	Phosphofructokinase
PGD	Phosphogluconate dehydrogenase
PGI	Phosphoglucose isomerase
PGK	Phosphoglycerate kinase
PPS	Phosphoenolpyruvate synthetase
PQO	Pyruvate:quinone oxidoreductase
PTA	Phosphotransacetylase
PTS	Phosphotransferase system
PYK	Pyruvate kinase
RPE	Ribulose 5-phosphate 3-epimerase
RPI	Ribulose 5-phosphate isomerase
SCS	Succinyl coenzyme A synthetase
SQO	Succinate:quinone oxidoreductase
TAL	Transaldolase
TKT1	Transketolase 1
TKT2	Transketolase 2
TPI	Triosephosphate isomerase

	Quantities	
$\hat{\Pi}_{\text{CO}_2, \text{tot}}$	experimentally determined (total) CO_2 -production rate	mmol h^{-1}
μ	specific growth rate	h^{-1}
ω_C	mass fraction of carbon in CDW	g C (g CDW)^{-1}
$\Phi_{\text{CO}_2}^\alpha$	volume fraction of CO_2 in the inlet air	vol%

$\Phi_{O_2}^\alpha$	volume fraction of oxygen in the inlet air	vol%
$\Phi_{CO_2}^\omega$	volume fraction of CO ₂ in the exhaust gas	vol%
$\Phi_{O_2}^\omega$	volume fraction of oxygen in the exhaust gas	vol%
$\Pi_{CO_2,spec}$	specific CO ₂ -production rate	mmol (g h) ⁻¹
$\Pi_{CO_2,tot}$	(total) CO ₂ -production rate	mmol h ⁻¹
$\Pi_{CO_2,vol}$	volumetric CO ₂ -production rate	mmol (l h) ⁻¹
$\Pi_{Glc,spe}$	specific glucose uptake rate	mmol (g h) ⁻¹
$\Pi_{Glc,tot}$	(total) glucose uptake rate	mmol h ⁻¹
$\Pi_{Glc,vol}$	volumetric glucose uptake rate	mmol (l h) ⁻¹
ρ	density	g l ⁻¹
Θ	instantaneous carbon balance	-
c_{CS}	concentration of a metabolite in the culture supernatant sample	μM
c_{Ext}	concentration of a metabolite in the extract sample	μM
$c_{Glc,0}$	initial concentration of glucose	mM
c_{intra}	intracellular concentration of metabolite	mM
c_{QS}	concentration of a metabolite in the quenching supernatant sample	μM
F_{air}	gassing rate	m ³ h ⁻¹
k^+	lumped forward rate constant for bicarbonate formation	h ⁻¹
k^-	lumped backward rate constant for bicarbonate formation	h ⁻¹
K_{H,CO_2}	Henry's law volatility constant	l Pa mol ⁻¹
$k_L a$	overall mass transfer coefficient	h ⁻¹
M_{Carbon}	molecular weight of carbon	g mol ⁻¹
n_{CS}	amount of a metabolite in the total samples culture supernatant	nmol
$n_{Leakage}$	amount of a metabolite leaked into the quenching solution	nmol
n_{QS}	amount of a metabolite in the total quenching supernatant	nmol
P_{ges}	pressure	Pa
pCO_2	partial pressure of carbon dioxide	Pa
V_{air}	gas volume in bioreactor vessel	m ³
V_{Ext}	volume of extraction solvent	ml
V_{liq}	liquid volume of bioreactor	l
V_{Quench}	volume of quenching solvent	ml
V_{Sample}	volume of culture sample	ml
X	biomass concentration	μl ml ⁻¹ or g l ⁻¹
X_0	initial biomass concentration	μl ml ⁻¹ or g l ⁻¹
$[CO_{2(aq)}]$	concentration of dissolved CO ₂	mol l ⁻¹
$[HCO_3^-]$	concentration of hydrogen bicarbonate	mol l ⁻¹
m/z	mass of an ion divided by its charge number and the unified atomic mass unit	-
bmACCOA	biomass requirement of metabolite pool ACCOA	mmol (g h) ⁻¹

BV	biovolume	μl
CDW	cell dry weight	g
DO	dissolved oxygen	%
OD ₆₀₀	optical density at 600 nm	-
R	universal gas constant	J (mol K) ⁻¹
rpm	revolutions per minute	min ⁻¹
T	temperature	K
vvm	inlet air gas flow per cultivation volume	l (l min) ⁻¹

Symbols and Notation

$\boldsymbol{\beta}$	vector of regression coefficients
$\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f}$	sensitivity matrix
$\hat{\boldsymbol{\beta}}$	estimated vector of regression coefficients
$\hat{\boldsymbol{\lambda}}$	estimated vector of response variables
$\hat{\phi}$	estimated overdispersion parameter
$\hat{\sigma}_{wj}$	estimated standard deviation of measured extracellular flux $\hat{w}_j$
$\hat{\sigma}_{yi}$	estimated standard deviation of measured enrichment $\hat{y}_i$
$\hat{w}_j$	measured extracellular flux j
$\hat{y}_i$	measured enrichment i
$\kappa^{F,C}$	number of carbon atoms in a fragment ion
$\kappa^{F,N}$	number of nitrogen atoms in a fragment ion
$\kappa^{P,C}$	number of carbon atoms in a precursor ion
$\kappa^{P,N}$	number of nitrogen atoms in a precursor ion
$\lambda^{F,C}$	number of ¹³ C-labeled carbon atoms in a fragment ion
$\lambda^{F,N}$	number of ¹⁵ N-labeled nitrogen atoms in a fragment ion
$\lambda^{P,C}$	number of ¹³ C-labeled carbon atoms in a precursor ion
$\lambda^{P,N}$	number of ¹⁵ N-labeled nitrogen atoms in a precursor ion
$\mathbf{g}(\mathbf{x})$	function mapping the weight-ordered cumomer vector onto simulated measurement values
$\mathbf{h}(\mathbf{c}_M)$	function mapping the metabolite concentration vector onto measured metabolite concentrations
$\mathbf{M}_v$	flux measurement matrix
$\mathbf{N}$	stoichiometric matrix
$\mathbf{X}$	design matrix
$\mathbf{x}$	weight-ordered cumomer vector of intracellular metabolites
$\mathbf{x}_{\text{inp}}$	cumomer vector of substrate
$\mathbf{z}$	vector of measured pool sizes
$\Phi_{f(p)}$	effect of the f^{th} fragment of peptide p
Π_p	effect of the p^{th} peptide

$\mathbf{C}_m^f$	matrix mapping the isotopomers of metabolite m onto tandem mass isotopomers of fragment f
$\mathbf{v}$	flux vector
$\mathbf{v}_f$	free flux vector
$Cov(\hat{\boldsymbol{\beta}})$	covariance matrix of the estimated vector of regression parameters
F_t^{-1}	inverse function of the cumulative distribution function of the t-distribution
I_i	effect of the i^{th} isotope
k	total mass shift of a metabolite, i.e. total number of labeled atoms
R_r	effect of the r^{th} replicate
T_t	effect of the t^{th} treatment
w_j	computed extracellular flux j
y_i	computed enrichment i
rDOF	remaining degrees of freedom
	Units
$^{\circ}\text{C}$	degrees Celsius
C	Coulomb
g	gram
h	hour
J	Joule
K	Kelvin
l	liter
M	mole per liter
m	meter
min	minute
mol	mole
Pa	Pascal

Introduction

1 Motivation

The anaplerotic node represents a circular reaction network around the metabolites phosphoenolpyruvate (PEP), pyruvate (PYR), oxaloacetate (OAA) and malate (MAL), which comprises specific carboxylation and decarboxylation reactions [158]. These reactions establish a link between glycolysis and the tricarboxylic acid cycle (TCA cycle) and therefore their mode of operation greatly affects the overall carbon flux into biosynthetic products emanating from the TCA cycle, like L-lysine (LYS) or L-glutamate (GLU). The annual production of these two products amounts to more than one million tons and more than two million tons, respectively, and is primarily accounted for by the biotechnological usage of engineered strains of *Corynebacterium glutamicum* [17]. Moreover, *C. glutamicum* has already been proven to possess exceptional capabilities as production host for other amino acids like L-phenylalanine (PHE) [208], L-valine (VAL) [138], L-leucine (LEU) [184] and L-isoleucine (ISOLEU) [185]. All these products derive from one or two of the above mentioned metabolites at the anaplerotic node. Therefore, the interplay of reactions at the anaplerotic node influences the availability of the respective precursor metabolites and consequently the respective yield. Figure 1.1 shows an overview of the anaplerotic node in *C. glutamicum*.

C. glutamicum distinguishes itself in possessing the genetic endowment for five enzymatic activities at the anaplerotic node, all of which are reported to show *in vitro* activity when assayed in or extracted from glucose grown cells of *C. glutamicum* strains [22, 33, 178, 97]. Anaplerotic carboxylation arises from the activities of (1) Phosphoenolpyruvate carboxylase (PEPCx, EC 4.1.1.31) and (2) Pyruvate carboxylase (PCx, EC 6.4.1.1). On the other hand, C3-directed decarboxylation is catalyzed by the enzymes (3) Phosphoenolpyruvate carboxykinase (PEPCk, EC 4.1.1.32) and (4) Oxaloacetate decarboxylase (ODx, EC 4.1.1.3). Furthermore, (5) Malic enzyme (ME, EC 1.1.1.40) catalyzes a bidirectional reaction between MAL and PYR [207, 63, 33]. Therefore, six different reactions catalyzed by five enzymes constitute the reaction network of *C. glutamicum* at the anaplerotic node. There is still little understanding of the regulation and essentiality of individual anaplerotic reactions when multiple carboxylation reactions are present. However, due to the importance of these reactions for the above mentioned products, a better understanding is clearly needed.

To study single *in vivo* anaplerotic activities, only the application of ^{13}C -Metabolic Flux Analysis (^{13}C -MFA) may give insight into their activity since different reactions catalyze stoichiometrically identical reactions. Early ^{13}C -MFA studies of the anaplerotic node in *C. glutamicum* not only yielded conflicting results on anaplerotic fluxes [144, 98], but also made non-consistent assumptions on the enzymatic activity of particular reactions. Some resorted to network simplifications like lumping of MAL and OAA metabolite pools [117, 144], while others neglected decarboxylation activities [140]. More recent studies devoted to flux estimation in *C. glutamicum* included the carbon atom transitions of all enzymatic activities at the anaplerotic node, but repeatedly obtained non-identifiable

2 Outline of this thesis

First, an identifiability analysis is performed to answer the question of whether all anaplerotic fluxes in *C. glutamicum* can be inferred at all from stationary ^{13}C -label enrichment data (Figure 1.2). This analysis leads to a set of deletion mutants, for which the flux distribution can be determined. Other mutants will be included into the strain library for the purpose of consistency checking. With the aim of gaining as deep an insight into the metabolism and adaptation of individual mutants as analytically possible, each strain is thoroughly phenotyped by extracellular and intracellular measurements. In particular, metabolome, proteome and fluxome studies are performed for selected mutant strains (middle section of Figure 1.2). The comparative proteome analysis harvests a novel regression approach to arrive at accurate and precise fold changes for a large part of the *C. glutamicum* proteome.

The inconclusive and inconsistently estimated fluxes from the ^{13}C -MFA study motivate the development of a new QqTOF-based ^{13}C -profiling method to which the second part of the work is devoted. This includes not only the establishment of an acquisition method, but also a thorough interpretation of liquid chromatography-electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) product ion spectra. On the way, it is investigated how the product ion spectra of amino acids can be used to achieve a separation of ^{13}C - and ^{15}N -isotopes.

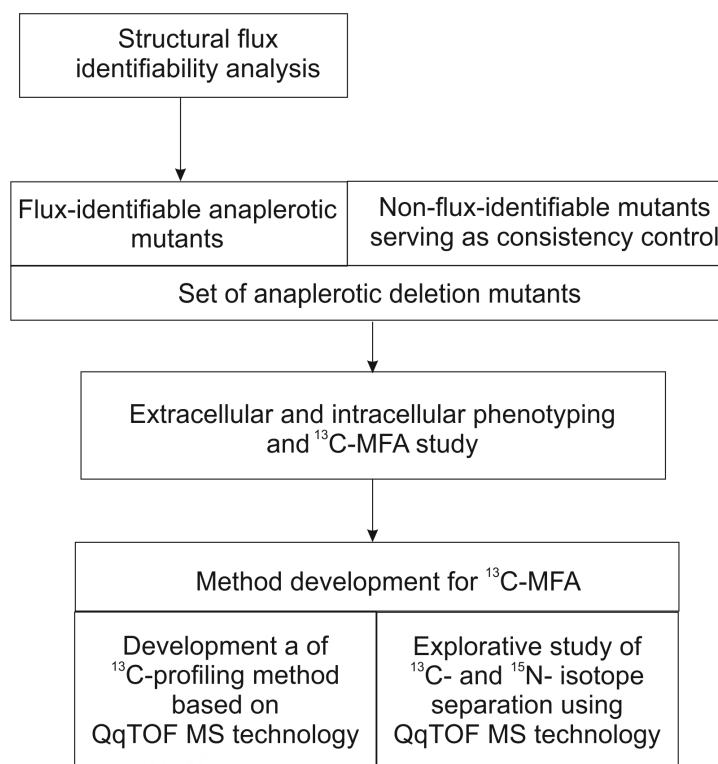


Figure 1.2: Outline of this thesis

Background

3 Anaplerotic Reactions and their Role in the Central Carbon Metabolism

The central carbon metabolism is an evolutionary conserved set of reactions which fulfills three major functions: Firstly, its constituent reactions produce reducing equivalents in the form of nicotinamide adenine dinucleotide (NADH), nicotinamide adenine dinucleotide phosphate (NADPH), and flavin adenine dinucleotide (FADH₂) or equivalently Menaquinol (MQH₂) as in *C. glutamicum*. Secondly, triphosphorylated ribonucleosides (adenosine 5-triphosphate, ATP; guanosine 5-triphosphate, GTP; etc.) and desoxyribonucleosides are generated either directly through substrate level phosphorylation or by means of oxidation of NADH, FADH₂, or MQH₂ during oxidative phosphorylation. Thirdly, these reactions provide a set of 12 precursor metabolites which are polymerized using NADPH and triphosphorylated ribonucleosides to yield biological macromolecules.

The central carbon metabolism with the peculiarities of *C. glutamicum* is depicted in Figure 2.1 in which colors highlight its main partitions, namely the Embden-Meyerhof-Parnas pathway (EMP pathway), the pentose phosphate pathway (PP pathway) and the TCA cycle. These pathways are connected by linker reactions. For instance, the activities of Transaldolase (TAL), Transketolase 1 (TKT1) and Transketolase 2 (TKT2) connect the PP pathway to the EMP pathway (see Figure 2.1). The latter is wired to the TCA cycle by two classes of reactions, the first of which comprises the above mentioned reactions at the anaplerotic node (also termed PEP-PYR-OAA node [158]). The second link between the EMP pathway and the TCA cycle is established by decarboxylations which act on PYR and coenzyme A (COA) to yield the thioester bond of acetyl coenzyme A (ACCOA). The latter is assimilated into the TCA cycle by Citrate synthase (CS) and subsequently oxidized to two equivalents of CO₂ in net terms by means of two consecutive decarboxylations. An alternative route to these decarboxylations is catalyzed by the enzymes of the glyoxylate shunt (GLX shunt), i.e. Isocitrate lyase (ICL) and Malate synthase (MS), which circumvent the reactions of Isocitrate dehydrogenase (ICD), the Oxoglutarate dehydrogenase complex (ODHC) and Succinyl coenzyme A synthetase (SCS).

Figure 2.1 also highlights the reactions which withdraw precursor metabolites for the synthesis of macromolecules. Every increase in biological matter composed of macromolecules requires an organism-specific equivalent of these precursor metabolites per equivalent of biomass formed and therefore growth as such will deplete these metabolites unless they are replenished. The action of replenishment is meant by the literal meaning of the word anaplerosis which is a compound word assembled from the Greek prefix ἀνά (ana-, "again, upward"), the Greek adjective πλήρης (pleres, "full"), and the Greek suffix -ωσις (-osis, "action") [191]. But exactly what is meant to be replenished is not implicit in the literal meaning of this term. Neidhardt et al. [129] invokes the term within the context of replenishing the set of 12 precursors metabolites:

"The interconnecting, reversing, and bypassing reactions are termed anapleurotic reactions; they replenish the pools of precursor metabolites drained by biosynthesis."

3. ANAPLEROTIC REACTIONS AND THEIR ROLE IN THE CENTRAL CARBON METABOLISM

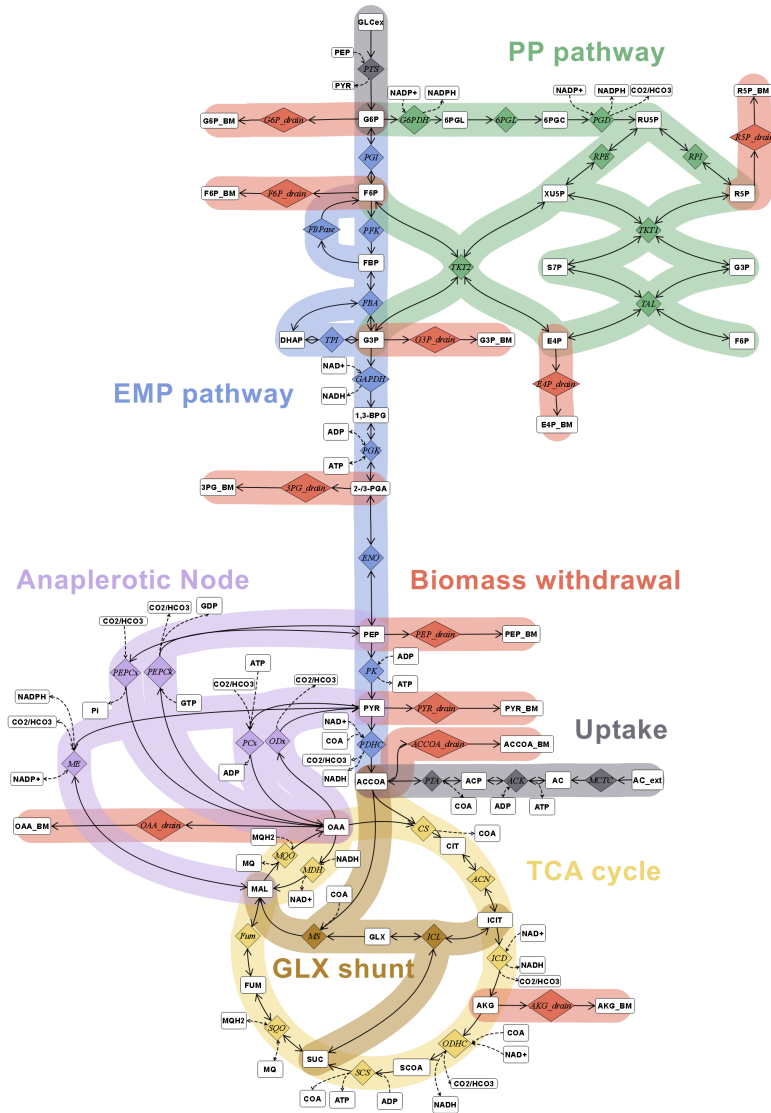


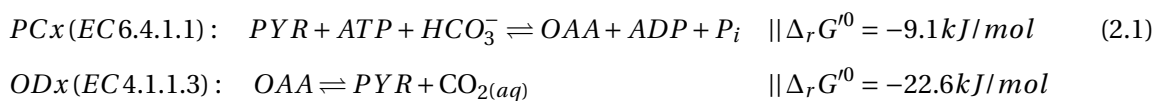
Figure 2.1: Schematic of the central carbon metabolism of *C. glutamicum* with partitions differentiated by colors. Abbreviations: PTS, Phosphotransferase system; G6PDH, Glucose 6-phosphate dehydrogenase; 6PGL, 6-Phosphogluconolactonase; PGD, Phosphogluconate dehydrogenase; RPE, Ribulose 5-phosphate 3-epimerase; RPI, Ribulose 5-phosphate isomerase; TKT1, Transketolase 1; TKT2, Transketolase 2; TAL, Transaldolase; PGI, Phosphoglucose isomerase; PFK, Phosphofructokinase; FBPase, Fructose 1,6-bisphosphatase; FBA, Fructose-bisphosphate aldolase; TPI, Triosephosphate isomerase; GAPDH, Glyceraldehyde phosphate dehydrogenase; PGK, Phosphoglycerate kinase; CS, Citrate synthase; ACN, Aconitase; ICD, Isocitrate dehydrogenase; ODHC, Oxoglutarate dehydrogenase complex; SCS, Succinyl CoA synthetase; SQO, Succinate:quinone oxidoreductase; Fum, Fumarate; ASAT, Aspartate transaminase; LDH, Lactate dehydrogenase; ENO, Enolase; ICL, Isocitrate lyase; MDH, Malate dehydrogenase; ME, Malic enzyme; MQO, Malate:quinone oxidoreductase; MS, Malate synthase; ODx, Oxaloacetate decarboxylase; PCx, Pyruvate carboxylase; PDHC, Pyruvate dehydrogenase complex; PEPCK, Phosphoenolpyruvate carboxykinase; PEPCK, Phosphoenolpyruvate carboxylase; PTA, Phosphotransacetylase; ACK, Acetate kinase. Since ACCOA is used during methionine biosynthesis in *C. glutamicum* rather than succinyl coenzyme A (SCOA) as in *E. coli* [107], the set of precursor metabolites actually comprises eleven intermediates in *C. glutamicum* and not 12 as in *E. coli*.

Obviously, many precursor metabolites are produced by virtue of being intermediates of energy generating pathways. For instance, when the carbon source is D-glucose, the reaction sequences of the EMP and PP pathways serve all three of the above mentioned purposes. In contradistinction, when the carbon source is a TCA cycle intermediate, the carbon flow through the EMP pathway needs to be reversed in direction to yield precursor metabolites of the EMP and PP pathways. Obviously, reducing equivalents and ATP need to be spent for this reversal of directionality of the EMP and PP pathway reactions. Amongst these reactions are TAL, TKT1, and TKT2 which Neidhardt et al. [129] classify as anaplerotic under the latter condition, but do not so under the former one. Hence, this definition implies the condition that a reaction may not serve any other purpose but to replenish biomass precursors for it to qualify as anaplerotic reaction.

Whereas the definition of Neidhardt et al. [129] links anaplerosis to the set of 12 precursor metabolites, the term was first coined in a more narrow sense of replenishing the intermediates of the TCA cycle by Kornberg [103]. During one turn of this metabolic cycle one equivalent ACCOA entering the cycle is oxidized to two equivalents carbon dioxide, four reducing equivalents and one equivalent ATP (see Figure 2.1). Hence, every carbon atom assimilated into the TCA cycle in the form of ACCOA is lost as CO₂. Consequently, every reaction consuming some of its intermediates will drain this cycle and must be matched by a refueling reaction for the cycle to continue. Referring to this fact, Kornberg [103] suggested the following definition:

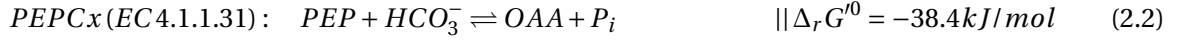
"[...] some route ancillary to that cycle (the TCA cycle) must operate to ensure that the intermediates of the cycle, which are drained away during growth, are replenished. [...] these obligatory routes of replenishment [...] may be generally described as anaplerotic sequences."

One reaction able to catalyze such an anaplerotic sequence in the sense of this definition is PCx which assimilates CO₂ into the metabolism by means of condensation with PYR in an ATP-dependent manner to yield OAA. However, its activity does not guarantee that a net supply of biomass precursors is available. For as long as decarboxylating reactions are active, only the net rate at which TCA intermediates are formed is available for TCA drainage. Therefore, the decarboxylation reaction catalyzed by ODx needs to be considered as well for it converts OAA to PYR. The balanced reactions of this pair are given as:



Moreover, the carboxylation reaction of PEPcX also yields OAA through condensation with bicarbonate to PEP, while releasing one phosphate. Its activity is counteracted by the reaction of PEPcK which catalyzes the decarboxylation of OAA to yield PEP. Both reactions are given as:

3. ANAPLEROTIC REACTIONS AND THEIR ROLE IN THE CENTRAL CARBON METABOLISM



Apart from the above mentioned reaction pairs, ME catalyzes a conversion between PYR and MAL that is coupled to $NADP^+$ according to the following reaction:



To the right of each equation, the standard Gibbs energy of reaction is shown which was estimated using the eQuilibrator tool [53]¹. Surprisingly, the reactions of PEPCk and ME actually show positive standard Gibbs free energies of reaction for the assumed direction. Therefore, all assumptions on reversibility and net directionality are eschewed for the time being to render the following discussion as general as possible. Further below, experimental evidence is quoted which allows to make assumptions on the unidirectionality of some reactions in *C. glutamicum*.

Under glycolytic conditions it is quite intuitive to see that the "anaplerotic sequence" in the sense of Kornberg's definition [103] consists of one step, i.e. the aforementioned carboxylations. However, when more complex sequences operate to effect a TCA replenishment (see below), this definition is not clear as to whether all enzymes of the sequence shall be denoted anaplerotic reactions or just a subset. This shortcoming and the inconsistencies between existing definitions motivate the attempt to provide a consistent and unifying definition of an anaplerotic reaction. This definition resides on the requirement that a reaction has to result in a net replenishment of carbon atoms of the TCA cycle for it to qualify as anaplerotic reaction. In particular, such a reaction has to conform to the following definition:

Definition 3.1. An anaplerotic reaction is any intracellular reaction whose net flux contributes positively to the carbon balance of the TCA cycle. Every intracellular reaction whose netflux drains carbon atoms of it is a cataplerotic reaction.

To deduce which reactions are covered by the above definition under any given physiological state, it is instructive to set up this TCA cycle carbon balance. To this end, it seems logical to consider only the carbon atoms any reaction contributes or removes from the TCA cycle. Consequently, it is appropriate to model the reactions of each carboxylation/decarboxylation pair as one forward and one backward flux of the same reaction. Although both reactions of the pairs of equations (2.1) and (2.2) differ in the specific co-factors, they are identical from the viewpoint of a carbon balance for they are the same with respect to the pools of balanced carbon atoms. Therefore, we deal with three

¹The web interface was accessed on 27 September 2018.

reversible carboxylations that act on PYR or PEP and yield either OAA or MAL as TCA cycle intermediates. To set up the carbon balance of the entire TCA cycle, one has to consider all its intermediates and set up individual carbon balances around these metabolites. To formulate these balances, the most general case shall be considered where all reactions producing or consuming TCA intermediates according to Figure 2.1 are active. Then, a given physiological state, under which some reactions are not active, represents a simplification of the most general case. In these balances, the carboxylation reactions/decarboxylation pairs are considered as forward in the direction of OAA and MAL, whereas the reactions of *OAA_drain* and *AKG_drain* are defined as forward in the direction of TCA cycle outflow. Given these definitions, the following balances hold:

$$\begin{aligned}
 \text{CIT :} \quad & 6 \cdot v_{\text{CS}}^{\text{net}} - 6 \cdot v_{\text{ACN}}^{\text{net}} = 0 \\
 \text{ICIT :} \quad & 6 \cdot v_{\text{ACN}}^{\text{net}} - 6 \cdot v_{\text{ICL}}^{\text{net}} - 6 \cdot v_{\text{ICD}}^{\text{net}} = 0 \\
 \text{GLX :} \quad & 2 \cdot v_{\text{ICL}}^{\text{net}} - 2 \cdot v_{\text{MS}}^{\text{net}} = 0 \\
 \alpha\text{KG :} \quad & 5 \cdot v_{\text{ICD}}^{\text{net}} - 5 \cdot v_{\text{ODHC}}^{\text{net}} - 5 \cdot v_{\text{AKG_drain}}^{\text{net}} = 0 \\
 \text{SCOA :} \quad & 4 \cdot v_{\text{ODHC}}^{\text{net}} - 4 \cdot v_{\text{SCS}}^{\text{net}} = 0 \\
 \text{SUC :} \quad & 4 \cdot v_{\text{SCS}}^{\text{net}} + 4 \cdot v_{\text{ICL}}^{\text{net}} - 4 \cdot v_{\text{SDH}}^{\text{net}} = 0 \\
 \text{FUM :} \quad & 4 \cdot v_{\text{SDH}}^{\text{net}} - 4 \cdot v_{\text{FUM}}^{\text{net}} = 0 \\
 \text{MAL :} \quad & 4 \cdot v_{\text{FUM}}^{\text{net}} + 4 \cdot v_{\text{ME}}^{\text{net}} + 4 \cdot v_{\text{MS}}^{\text{net}} - 4 \cdot v_{\text{MDH_MQO}}^{\text{net}} = 0 \\
 \text{OAA :} \quad & 4 \cdot v_{\text{MDH_MQO}}^{\text{net}} + 4 \cdot v_{\text{PCx_ODx}}^{\text{net}} + 4 \cdot v_{\text{PEPCx_PEPCk}}^{\text{net}} - 4 \cdot v_{\text{OAA_drain}}^{\text{net}} - 4 \cdot v_{\text{CS}}^{\text{net}} = 0
 \end{aligned}$$

$$2 \cdot v_{\text{CS}}^{\text{net}} - v_{\text{ICD}}^{\text{net}} - v_{\text{ODHC}}^{\text{net}} - 5 \cdot v_{\text{AKG_drain}}^{\text{net}} - 4 \cdot v_{\text{OAA_drain}}^{\text{net}} + 4 \cdot (v_{\text{PCx_ODx}}^{\text{net}} + v_{\text{ME}}^{\text{net}} + v_{\text{PEPCx_PEPCk}}^{\text{net}}) + 2 \cdot v_{\text{MS}}^{\text{net}} = 0 \quad (2.4)$$

After summation and cancellation of terms, the total carbon balance of the TCA cycle of equation (2.4) arises. This equation clearly represents the fact that both carbon atoms which are assimilated into the TCA cycle are subsequently lost through the activities of ICD and ODHC. Depending on the physiological state, equation (2.4) may be simplified which will be illustrated in the following using three physiological states. The first physiological state is depicted in Figure 2.2 which shows a restriction of the larger model of Figure 2.1 and depicts the TCA cycle operation when metabolism is glycolytic and aerobic. Under this condition the GLX shunt is inactive, indicated by missing coloration of its two enzymes ICL and MS. Furthermore, the full sequence of the TCA cycle operates such that the following linear dependencies hold: $v_{\text{ICD}}^{\text{net}} = v_{\text{CS}}^{\text{net}}$ and $v_{\text{ODHC}}^{\text{net}} = v_{\text{CS}}^{\text{net}} - v_{\text{AKG_drain}}^{\text{net}}$. Furthermore, the rate of the MS-catalyzed reaction is zero, i.e. $v_{\text{MS}}^{\text{net}} = 0$. Through substitution of these constraints into equation (2.4), cancellation, and collection of terms one obtains:

$$-v_{\text{AKG_drain}}^{\text{net}} - v_{\text{OAA_drain}}^{\text{net}} + v_{\text{PCx_ODx}}^{\text{net}} + v_{\text{ME}}^{\text{net}} + v_{\text{PEPCx_PEPCk}}^{\text{net}} = 0 \quad (2.5)$$

Equation (2.5) shows that the outflow of the TCA cycle from OAA and α -ketoglutarate (α KG) must

3. ANAPLEROTIC REACTIONS AND THEIR ROLE IN THE CENTRAL CARBON METABOLISM

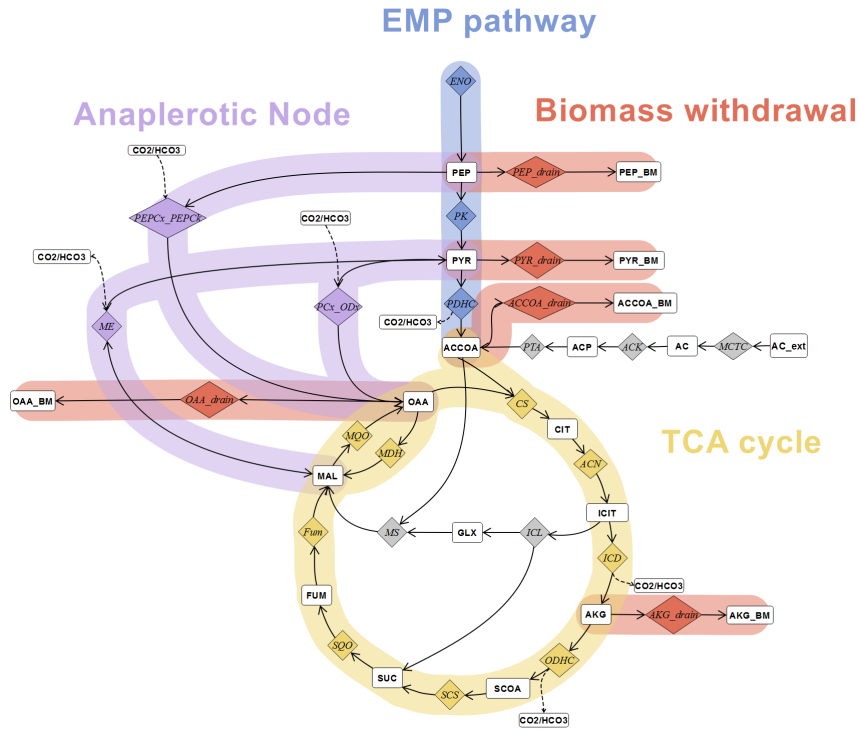


Figure 2.2: TCA operation under aerobic conditions when the carbon source is D-glucose. The inactivity of the GLX shunt is indicated by the missing color.

be matched by the sum of the net carboxylation activity of the reactions pairs of PCx_ODx , ME , and $PEPCx_PEPCK$. When any of these reactions has a positive net flow, it qualifies as anaplerotic reaction for they are defined as forward in the direction of carboxylation. Under this particular condition, the definition used in this thesis agrees with the one of Kornberg [103] who defines the carboxylations as anaplerotic reactions under glycolytic conditions.

By contrast, when the carbon source is acetate the assimilation of the carbon source does not yield a substrate to any of the above carboxylation reactions. Instead, assimilation of acetate yields ACCOA as central carbon metabolism intermediate. Given the fate of the carbon atoms of ACCOA during TCA cycle operation, another sequence must exist to replenish TCA intermediates. Back in 1957, Kornberg and Krebs [105] correctly surmised that acetate assimilation relies on the GLX shunt for TCA cycle replenishment. This bypass reaction spares the carbon backbone of ICIT from decarboxylation by converting it to two products, i.e. succinate (SUC) and GLX, the latter of which is condensed with ACCOA to yield MAL (see Figure 2.3). Although these two reactions are necessary, they do not represent the fully balanced GLX cycle. In fact, to effect the net synthesis of a C4 body the following reaction sequence is required which represents one turn of the GLX cycle [103]:

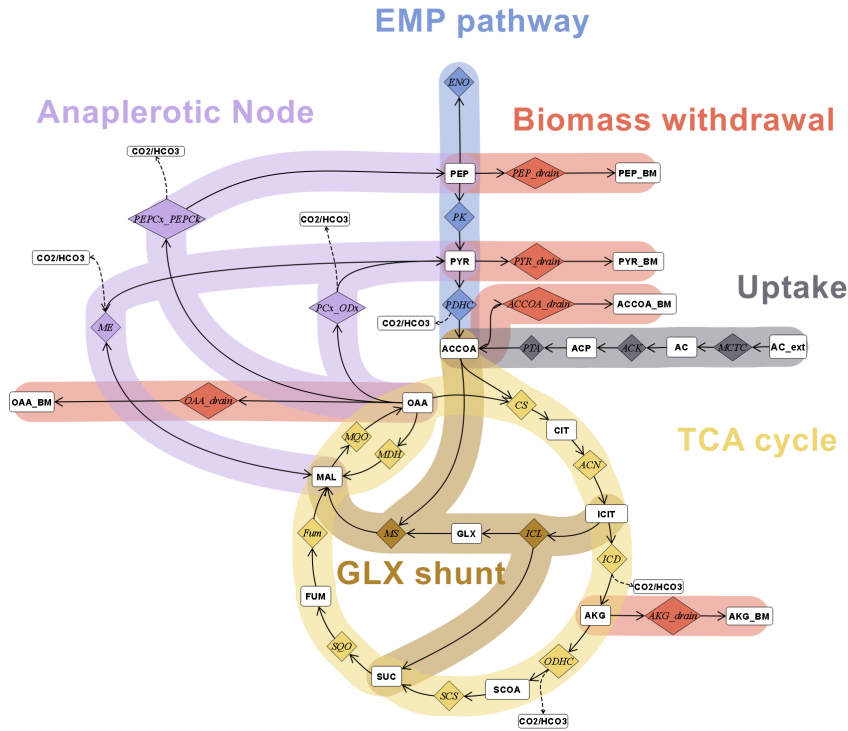
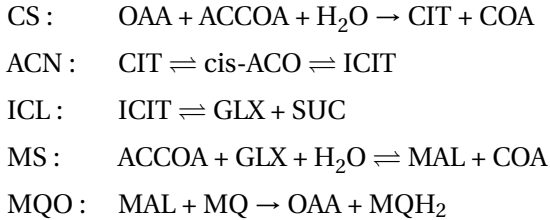


Figure 2.3: TCA operation under aerobic conditions when the carbon source is acetate. Apart from the TCA cycle and the anaplerotic node, the enzymes of the GLX shunt as well as those required for acetate assimilation are active.



$$2 \cdot \text{ACCOA} + 2\text{H}_2\text{O} + \text{MQ} \rightarrow \text{SUC} + 2 \cdot \text{COA} + \text{MQH}_2 \quad (2.6)$$

Apparently, for the net generation of one equivalent of SUC the reactions of CS, ACN and MQO are required as well. The question arises whether these reactions shall also be classified as anaplerotic by virtue of being essential to the function of the GLX shunt. To answer this question, we recur to the TCA cycle carbon balance of equation (2.4). Since the GLX shunt is now additionally active, it holds: $v_{\text{ICD}}^{\text{net}} = v_{\text{CS}}^{\text{net}} - v_{\text{ICL}}^{\text{net}}$ and $v_{\text{ODHC}}^{\text{net}} = v_{\text{CS}}^{\text{net}} - v_{\text{ICL}}^{\text{net}} - v_{\text{AKG_drain}}^{\text{net}}$. Moreover, the rates of ICL and MS must be equal, i.e. $v_{\text{ICL}}^{\text{net}} = v_{\text{MS}}^{\text{net}}$. Through substitution of these stoichiometric constraints, cancellation, and collection of terms one obtains:

$$-v_{\text{AKG_drain}}^{\text{net}} - v_{\text{OAA_drain}}^{\text{net}} + v_{\text{PCx_ODx}}^{\text{net}} + v_{\text{ME}}^{\text{net}} + v_{\text{PEPCx_PEPCk}}^{\text{net}} + v_{\text{MS}}^{\text{net}} \quad (2.7)$$

3. ANAPLEROTIC REACTIONS AND THEIR ROLE IN THE CENTRAL CARBON METABOLISM

When acetate is assimilated, the net flow of carboxylating reactions is negative for they account for the reversed EMP pathway flux to provide biomass precursor metabolite of the EMP and PP pathways. The net carbon balance of equation (2.7) clearly shows that this outflow from the TCA cycle and the biomass withdrawal have to be matched by the activity of MS. Although CS, ACN, and MQO form an essential part of the GLX cycle, equation (2.7) clearly shows that their reaction rate does not determine the TCA outflow as they cancel out in the balance. Instead, only MS specifically determines the net rate at which carbon atoms enter the TCA cycle, i.e. by condensation of ACCOA. Therefore, its rate directly yields the TCA cycle outflow under steady state conditions. Kornberg [103] defines each reaction of the GLX cycle sequence, i.e. CS, ACN, ICL, MS, and MQO as being of 'anaplerotic nature' for each reaction is required for the net synthesis of one equivalent of C4 body. In contradistinction, the definition of this thesis yields that MS is the only anaplerotic reaction for it alone determines the net rate at which carbon atoms enter the TCA cycle.

The last example concerns the growth of a microbial culture under microaerobic conditions. Under microaerobic conditions the terminal electron acceptor molecule oxygen is in short supply, which thermodynamically disfavors all reactions yielding reducing equivalents that need oxygen to be oxidized. On the other hand, all reactions consuming either MQH₂ or NADH are favored and in fact serve to regenerate oxidized NAD⁺ which is essential to the function of the EMP pathway. As a consequence, the net flow of the reactions of MDH, Fum, and SQO revert in direction to consume reducing equivalents (see Figure 2.4). This sequence proceeds to SUC for the reaction of SCS can be assumed to be irreversible. As the SUC pool needs to be balanced, it is exported and in fact represents a common culture supernatant metabolite in microaerobic fermentations [137]. The excretion is essentially a TCA cycle drain and therefore an additional TCA cycle outflow reaction needs to be added to the balance of equation (2.4), henceforth designated as *exc_SUC*. The activity of ODHC ceases completely for it produces one reducing equivalent. Though yielding one reducing equivalent alike, the reaction of ICD does not come to a halt for it provides the precursors of the glutamate family of amino acids (see Figure 2.4). Setting v_{ODHC}^{net} to zero and substituting v_{ICD}^{net} for v_{CS}^{net} , one is left with the following balance:

$$v_{CS}^{net} - 5 \cdot v_{AKG_drain}^{net} - 4 \cdot (v_{OAA_drain}^{net} + v_{exc_SUC}^{net}) + 4 \cdot (v_{PCx_ODx}^{net} + v_{ME}^{net} + v_{PEPCx_PEPCK}^{net}) = 0 \quad (2.8)$$

Evidently, the reaction of CS is now an anaplerotic reaction when the definition is rigorously applied for it assimilates one carbon atom into the oxidative branch of the TCA cycle for every equivalent of α KG that leaves the TCA cycle. On the other hand, the net activity of the carboxylating reactions provides the carbon backbones emanating from the reductive branch of the TCA cycle.

Heretofore, the focus has lied on the net flow a certain carbon balance reaction contributes to the TCA cycle. The net flow of such a carbon balance equation is essentially the outcome of a ¹³C-MFA study. When fluxes are identifiable from ¹³C-labeling data, the result of ¹³C-MFA is one estimate for

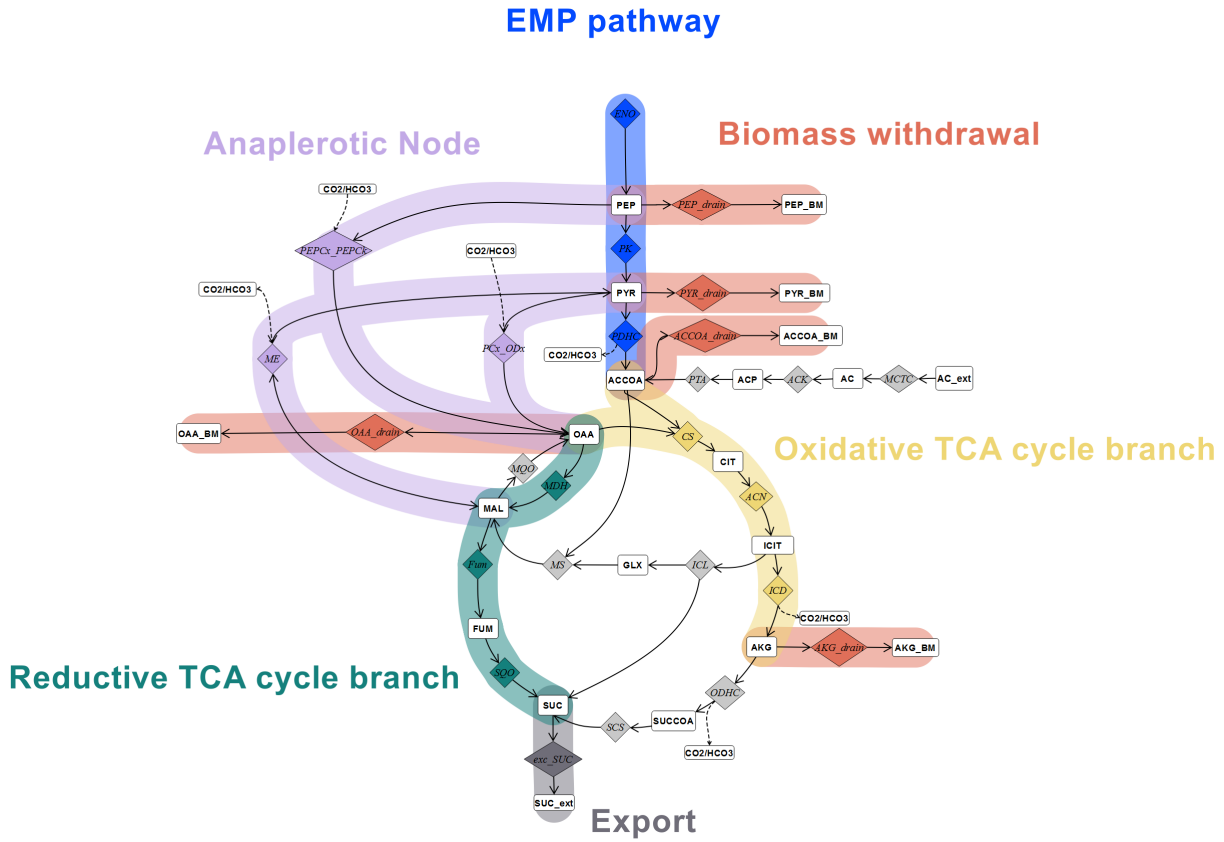


Figure 2.4: TCA operation under microaerobic conditions. The net directionality of the MQO/MDH reaction, Fum and SQO reverse and together they form the reductive branch of the TCA cycle. All active reactions are highlighted with color.

the backward and one for the forward flux for every modeled carbon balance equation, the difference of both being the net flow. When the activities of PEPCx and PEPCk are assumed to be reversible, there is no way to differentiate between the backward of ODx and the forward flow of PCx and vice versa. However, there is evidence to suggest that in *C. glutamicum* the reactions of some enzymes at the anaplerotic node are in fact irreversible. When each of the above mentioned reactions is irreversible, one reaction corresponds to either a forward or backward flux between the two metabolite pools it connects.

Firstly, a Δ PEPCk mutant is unable to grow on gluconeogenic substrates, which allows to deduce that PEPCx does not catalyze a decarboxylation [154]. Analogously, PCx does not appear to catalyze a decarboxylation either for a Δ ODx mutant shows no PYR forming activity when this activity is assayed *in vitro* [97]. On the other hand, Peters-Wendisch et al. [143] showed that a deletion mutant in PEPCx and PCx is not able to grow on glucose within 48 hours after inoculation. Apparently, PEPCk did not substitute for the missing carboxylation activity and neither did ODx. Therefore, both can be assumed to react unidirectionally. The same line of argument excludes ME as carboxylating enzyme. However, there are several studies which point to its carboxylation activity. Gourdon et al. [63] showed that ME purified from *C. glutamicum* is well able to catalyze carboxylation *in vitro*. Moreover, it could be

experimentally shown that a NADP⁺-dependent ME from *E. coli* can act as sole anaplerotic enzyme in *Saccharomyces cerevisiae* sustaining a growth rate of 0.06 h⁻¹ [207]. All this evidence warrants the consideration of one forward and backward reaction of ME, although under the particular condition quoted above it was not able to catalyze a carboxylation net flow.

4 LC-ESI-MS/MS instrumentation

4.1 Motivation

LC-ESI-MS/MS technology has developed into one of the indispensable workhorses of modern systems biology. Not surprisingly, a significant part of the data presented in this thesis were acquired using such devices. Therefore, this chapter first describes the basic working principles of MS/MS devices. The second part of this section is devoted to the three omics disciplines based on LC-ESI-MS/MS technology, i.e. metabolomics, proteomics and fluxomics as well as the way MS/MS technology is applied in each of these omics disciplines.

4.2 LC-ESI-QqQ MS

Established analytical LC-ESI-MS/MS workflows employ LC as first separation dimension, distinguishing analytes according to the strength of their interaction with the LC column material at the employed mobile phase conditions. The LC dimension mainly serves to separate the complexity of the biological matrix at hand and to enable proper ionization of target analytes in the ESI source, to which the LC eluent can be easily coupled. The LC eluent usually consists of a buffered aqueous phase and organic modifier and carries cellular metabolites which are not of interest, matrix ions as well as the target analytes. In proteomics the latter are peptides, whereas metabolomics and fluxomics target small molecules. ESI is particularly useful for both classes of analytes since non-volatile molecules such as highly polar metabolites and entire proteins can be easily ionized without need of derivatization [52]. Moreover, the ionization step itself imparts little energy in excess of the desolvation energy of small molecules [93] and therefore mostly yields intact molecular ions. In contradistinction, electron ionization (EI) as harsh ionization technique is known to fragment molecules upon ionization. The main drawback of ESI for the analysis of biological matrices, however, consists in its particular sensitivity to ion suppression effects. Components of the analyte matrix may reduce ionization efficiency of an analyte with which they co-elute; a phenomenon called matrix effect [119]. This places particular importance on proper chromatographic process design for a highly sensitive MS analysis.

When analytes are separated with LC methods working with pressures below 400 bar, they usually elute on a time scale of minutes (see time axis of the LC elution in Figure 2.5). This figure exemplarily shows an ion source in positive ionization mode. In this mode a positive voltage in the range of 3500 to 5500 V is applied to the electrospray capillary, making solvent molecules electrolyze. Water

molecules, for instance, electrolyze according to: $2 \text{H}_2\text{O} \rightarrow \text{O}_2 + 4 \text{H}^+ + 4 \text{e}^-$ [93]. With the electrons discharged over the electrospray capillary, a positive net charge is generated. The excess protons protonate other organic molecules resulting in even-electron ions of formula $[\text{M}+i\cdot\text{H}]^{i+}$, where i is a positive integer. Which ions are finally formed in the gas phase depends on the analyte and the experimental conditions. Metabolites usually have few protonation sites such that the $[\text{M}+\text{H}]^+$ ion is predominantly formed as gas phase species. Peptides, by contrast, have several protonation sites and usually generate multiply charged ions in the gas phase [94]. However, when other charge carriers such as Na^+ or Li^+ ions are present in the electrosprayed solution, non-specific adducts like $[\text{M}+i\cdot\text{Na}]^{i+}$ may additionally arise from the ionization [102]. In this case the total amount of analyte is spread across several ionized species of distinct mass, reducing sensitivity on any particular mass. This constitutes one way in which a particular analyte matrix can negatively affect the signal of a certain analyte ion.

Through the electric field applied between the electrospray needle and the counter-electrode, positively charged ions are drawn to the outer ending of the electrospray needle, forming a so-called Taylor cone at the end of it. Above a certain electric field strength, this Taylor cone is destabilized and disperses into solvent droplets in the μm -range, aided by a coaxial stream of nebulizer gas [93]. Due to the enrichment of positive charge in the Taylor cone, the generated droplets carry an excess of positive charge. High temperatures (in the range of 400-600°C), imparted by a stream of drying gas, further vaporize the solvent. With decreasing solvent content, the Coulomb repulsion of the ions bearing the same charge balances the surface tension of the droplet. At this so called Rayleigh limit, a Coulomb fissure occurs generating smaller droplets with higher surface tension. This process repeats itself several times till droplets are in the nm-range [14]. The process by which solvent-free gas-phase ions arise from these nano-droplets is not yet fully understood [102].

For low molecular weight compounds like metabolites, however, the *Ion Evaporation Model* has gained some popularity, not least because it has been supported by molecular dynamics simulations [4]. This model maintains that at the Rayleigh limit the nano-droplets eliminate cations clustered with solvent molecules rather than naked ions (see magnification in Figure 2.5) [102]. These gas-phase clusters drift along the electric field to the curtain plate which serves as counter-electrode to the electrospray needle. Through its aperture the ion-solvent clusters enter the vacuum interface region of the MS. The region between curtain plate and orifice is maintained at atmospheric pressure with pure nitrogen, flowing contrary to the movement of ions to avoid contamination of the vacuum interface with neutral species. The entering ion clusters collide with this nitrogen gas, leading to fragmentation of the clusters and release of the neutral solvent molecules. A voltage applied to the Orifice plate draws the ions further down the interface. This user-adjustable potential (called declustering potential, DP) is also used to tune the acceleration the ions experience against the nitrogen molecules in the orifice region [2]. Moderate values will contribute to the fragmentation of ion-solvent clusters, thus improving sensitivity. Higher voltages may lead to fissure of covalent bonds of the ion molecule. This process is called skimmer fragmentation. In principle, it is analogous to collision induced disso-

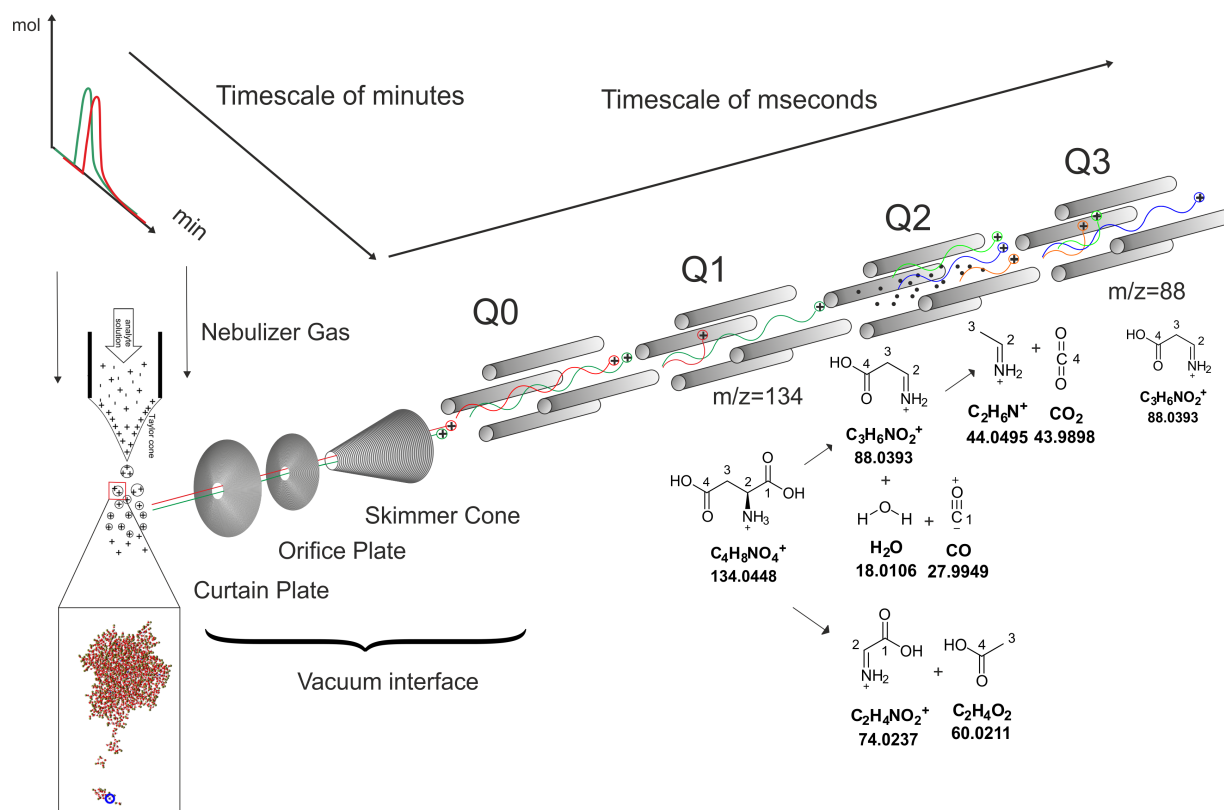


Figure 2.5: The schematic depicts an LC-ESI-QqQ MS analysis in positive ionization mode. The LC-eluent contains the target analyte (green color) and possible co-eluent (red color) as well as the mobile phase. The positively charged electrospray capillary induces the formation of a Taylor cone at its end which disperses into solvent droplets in the μm range, aided by an coaxial stream of nebulizer gas. High temperatures (in the range of 400 – 600 °C imparted by a stream of drying gas further vaporizes the solvent. At the Rayleigh limit (in nm-range) the droplets eliminate cations clustered with solvent molecules. These gas-phase clusters enter the vacuum interface region where they lose the solvent molecules through collisions with residual gas in the vacuum interface. Although the schematic refers to ESI-QqQ MS of the API 4000 series of AB Sciex, QqQ devices of other manufacturers like Thermo Scientific, Waters and Agilent are comprised of the same assembly groups such that a discussion of this particular device illustrates general manufacturer-independent principles of operation. Further explanations are given in the main text. The representation is not true to real dimensions. The representation of the ESI process was reprinted from Konermann et al. [102].

ciation without precursor ion selection [94].

After passing through the interface region, ions enter the Q0 quadrupole. This quadrupole is operated in radiofrequency (rf)-only mode and thus transmits all ions through its rods (designated by two ion trajectories in red and green in Figure 2.5). Its purpose is to separate the atmospheric pressure region of the ionization source from the high-vacuum region harboring the Q1, Q2 and Q3 quadrupoles. The Q0 is maintained at $8 \cdot 10^{-3}$ torr, whereas the high-vacuum region is operated at $1 \cdot 10^{-5}$ torr [1]. The first stage of mass analysis is performed in the Q1 quadrupole. Apart from a rf-field of alternating voltage, a direct current is applied to each rod of the quadrupole, while adjacent rods have reversed polarity. For one setting of direct current voltage and rf-amplitude, only ions of specific m/z value are able to describe a stable trajectory through the quadrupole (green ion is Figure

2.5). Ions of other m/z values have an unstable trajectory and collide with the quadrupole rod [113].

The next quadrupole is denoted Q2 or collision cell. It is maintained at a low pressure of inert collision gas, conventionally nitrogen. The Q2 quadrupole is flanked by lenses to reduce collision gas leakage into the high vacuum Q1 and Q3 quadrupole regions. A potential difference between Q0 and Q2 accelerates the precursor ion against the collision gas in Q2. This parameter is called collision energy (CE) and determines the kinetic energy to which an ion of given m/z is accelerated. The collisions of the ions with the nitrogen gas (designated by the black dots in quadrupole 2 in Figure 2.5) are inelastic such that part of the kinetic energy is converted into internal energy of the precursor ion [40], a process called activation. Higher values of CE impart higher kinetic energy and make more energy available for conversion to internal energy. The so obtained internal energy of the ion may be dissipated through vibrations of the vibrational degrees of the freedom of the molecule in question or may alternatively lead to covalent bond fissures. Since the activation step is several orders of magnitude faster than the fragmentation reaction of the activated precursor ion, the whole process is essentially an unimolecular reaction [168] (see ASP in Figure 2.5).

The observed spectrum of product ions depends on the kinetic energy of the precursor ion. Higher energy collisions favor energetically more expensive and frequent breakages, shifting the product ion spectrum to smaller masses. Regardless of the kinetic energy, the CID process produces mostly product ions that are even-electron species of type $[M-X+H]^+$, such that the eliminated losses X are even-electron neutral species. The typical neutral losses for even electron precursor ions of small molecules are carbon dioxide, carbon monoxide, water, ammonia, and small carboxylic acids. A common fragmentation for amino acids, for instance, involves the combined loss of water and carbon monoxide to yield an iminium ion [43]. This fragmentation is exemplarily depicted in Figure 2.5. Although these neutral losses are typical for CID fragmentation of ESI-ions of small molecules, they cannot be used to compute the product ion spectrum of metabolites. In fact, MS/MS spectra of small molecules are always empirical and their interpretation poses a veritable challenge [198].

Depending on the structure of the molecule, several *fragmentation channels* exists, i.e. mutually exclusive ways to fragment. An alternative channel to the iminium ion formation consists in the elimination of acetic acid, shown in the lower part of Figure 2.5. Furthermore, at higher nitrogen gas pressures the probability of multiple collisions along the length of the collision cell increases such that fragment ions from a previous fragmentation event may themselves be activated and fragment further [168]. This process is exemplarily depicted in Figure 2.5, in which the iminium ion at m/z 88.0393 fragments further to yield an ion at m/z 44.0495 through the elimination of CO_2 . Whereas the neutral eliminations in each reaction step are aspirated by the vacuum pump, the ions scattered by the collision events are refocused onto the ion axis by the rf-field prevalent in the Q2 quadrupole. Since the Q2 is operated in rf-mode only, all generated product ions are transmitted through this quadrupole to the second stage of mass analysis.

For the second stage of mass analysis of the formed product ions QqQ MS devices employ a third quadrupole. Since the setting of one specific direct current voltage and radio frequency ampli-

tude determines the m/z value that is transmitted through the quadrupole, these parameters need to be ramped through (scanned) in order to record the entire product ion spectrum. Therefore, quadrupole mass analyzers are called *scanning* mass spectrometers [48]. When a product ion spectrum is recorded in this way, most of the time would be spent acquiring signal on quadrupole transmission settings corresponding to m/z values for which no ions are present. Only a fraction of the time during which the analyte in question gets ionized would the detector actually record ions. Due to this property, quadrupoles are reckoned to afford low full spectrum sensitivity [169].

Not least because of this quadrupoles are most appropriately used for *Selected Reaction Monitoring* (SRM). SRM relies on transmitting one specific precursor mass in Q1 and one specific product ion mass in Q3. Figure 2.5 exemplarily shows the third quadrupole being set to $m/z=88$, transmitting the $m/z=88.0393$ ion only, whereas the other product ions collide with the quadrupole rod (orange and green product ion trajectories in Figure 2.5). Each such pair of Q1 and Q3 setting is called a *mass trace* (or transition) and corresponds to a precursor and product ion. It may be noted that a mass transition is defined by a set of two ions, essentially monitoring the reaction that yields the product ion from the precursor ion. Thus, the definition of a mass transition can be based on the exact masses of the chemical formulas of the precursor and product ion. The exact mass is the calculated mass for any elemental composition, indicated up to the fourth decimal position. Therefore, the following definition will be employed in this thesis:

Definition 4.1. A pair of precursor ion exact m/z and product ion exact m/z defines an exact mass transition.

This definition holds regardless of whether the MS may be able to actually measure the mass accurately. Quadrupoles are most commonly operated at unit resolution since the mass resolution these devices are able to achieve is low [113]. With unit resolution, all species within a ± 0.5 Da window of the target mass are co-isolated for one particular setting of Q1. Thus, when quadrupole technology is used, the precursor and product masses can only be indicated using integer numbers, therefore an additional definition is introduced:

Definition 4.2. A pair of precursor ion nominal m/z and product ion nominal m/z is called a nominal mass transition.

The signal on a nominal mass transition optimally corresponds to the one of the corresponding exact mass transition, i.e. the nominal mass trace monitors only the reaction for which it was derived. However, in complex samples the signal on one particular mass trace may be compromised by other co-isolated species such that the signal on a nominal mass transition does not correspond the one that would be obtained by acquiring an exact mass transition.

The sensitivity for any particular mass trace in SRM can be tuned by increasing the time spent transmitting this ion pair (dwell time). Depending on the particular MS device and metabolite, dwell times in the range of 5-150 ms are sufficient to achieve sub-micromolar sensitivity [114, 205]. When several target analytes are to be measured in one run, a greater number of mass transitions needs to

be monitored. These measurement setups are called *Multiple Reaction Monitoring* (MRM). In MRM assays the *cycle time* is the sum of the dwell times of all measured mass transitions plus time overhead per mass transition to clear the collision cells of ions (pause between mass ranges). After the pass of one cycle time, the signal on one particular mass trace can be acquired again. Since the timescale of the elution in conventional HPLC separations is in the range of several ten seconds to minutes, the total cycle time can assume a value in the range of seconds so as to obtain sufficient (≥ 10) data points across the HPLC peak on the mass trace corresponding to one metabolite. This places a *cycle time limit* on the number of mass traces that can be monitored in MRM assays using QqQ MS technology.

4.3 QqTOF MS technology

The cycle time limit inherent in scanning mass analyzers as described above can be circumvented by using a TOF mass analyzer for product ion detection. The hybrid of a quadrupole for precursor ion selection and TOF MS for product ion detection is called QqTOF MS. These hybrid devices have historically been used for identification purposes due to the high selectivity afforded by accurate mass measurements (see below). With regard to sensitivity (limited by the efficiency of ion transmission) and linear dynamic range (limited by the saturation of the time to digital converter) TOF mass spectrometers lagged far behind quadrupoles. However, in recent years newly developed QqTOF MS devices emerged with improved sensitivity and quantification capabilities similar to those of quadrupoles [8], which are now used routinely in quantitative studies [26, 61].

Contrary to a quadrupole mass analyzer, the TOF detector achieves mass separation in time instead of filtering ions through electric fields along a spatial dimension. For this purpose, the continuous ion beam from the first quadrupole stages is discretized into temporally discrete ion packs [67]. The discretization is performed by periodically accelerating a slice of the ion beam by means of an electric field; a process termed ion extraction. This electric field orthogonally accelerates the ion packs into a field free and evacuated region and imparts a kinetic energy equal to $q \cdot V$ on the ions, where V is the voltage of the accelerating field and q is the total charge of the ion in Coulomb. The latter is actually a product of the number of elementary charges z , i.e. the *charge number*, and the elementary charge e in Coulomb. Given identical initial kinetic energy, product ions of higher mass take longer to reach the detector due to smaller drift velocity than lower mass ions [113] (see product ion ordering in Figure 2.6). Therefore, the performed measurement is the time from acceleration to detection of an ion. This time is equal to the drift time t_D in the field free drift region, when the time in the acceleration region is neglected. For a linear TOF MS the drift time, in turn, relates to the mass to charge ratio $\frac{m}{q}$ of any ion according to equation (2.9) [67]. This equation presents the ideal case of an ion population without any variance in the initial conditions (initial position and initial velocity). The mass of the ion m in equation (2.9) is the mass in kg and q is the charge in Coulomb. The ratio of both quantities is not to be confounded by the so called m/z value which is a dimensionless quantity obtained by dividing the mass of the ion by the unified atomic mass unit ($1.667 \cdot 10^{-27}$ kg) and the charge number [128]. This quantity, though not equal to the mass to charge ratio $\frac{m}{q}$, is the commonly

used quantity to refer to a certain ion.

$$t_D = \frac{D}{\sqrt{2 \cdot V}} \sqrt{\frac{m}{q}} \quad (2.9)$$

Since all ions of the ion beam are accelerated and detected in time, TOF analyzers allow full-spectral acquisition of the entire product ion spectrum without the cycle time limitation inherent in full-spectral acquisition with quadrupoles described above. Although QqTOF MS-based acquisition methods acquire the entire product ion spectrum at once, the concept of a mass trace is still applicable. The intensity of the mass peak corresponding to the product ion in question is extracted from the TOF product ion spectra acquired over time and serves to reconstruct the *extracted ion chromatogram* (XIC).

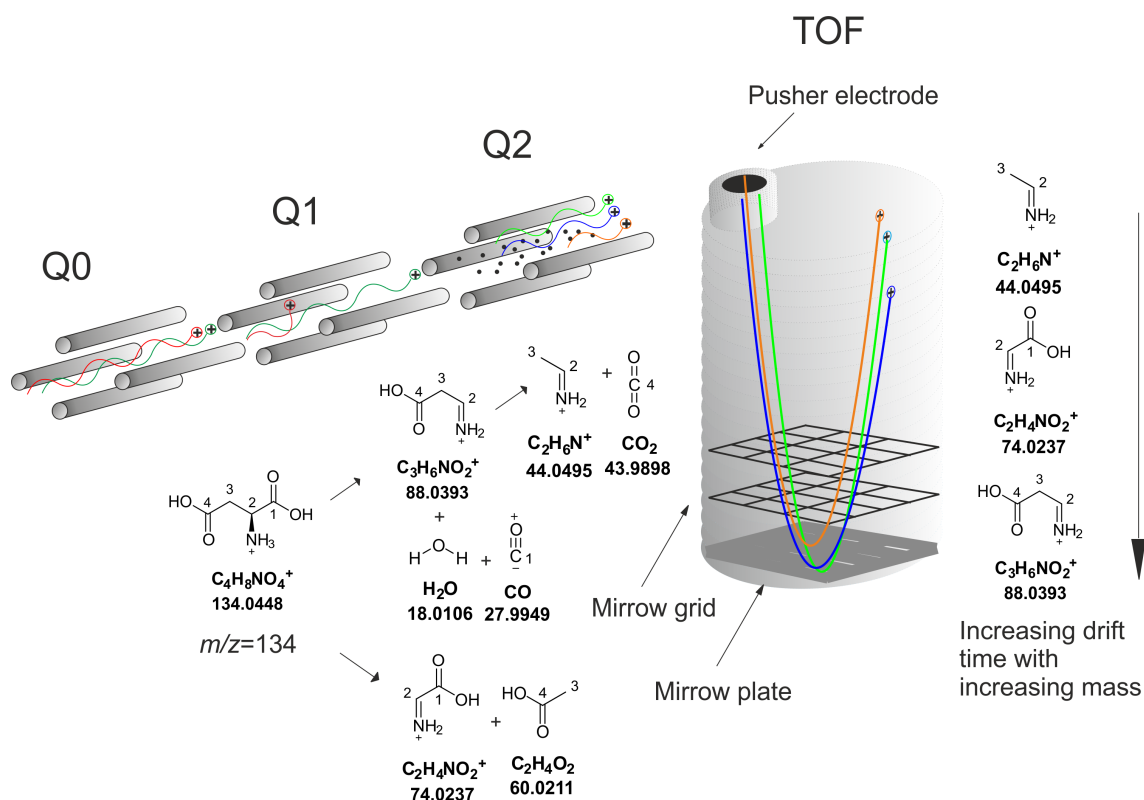


Figure 2.6: A schematic of an ESI-QqTOF MS device. The spectrum of product ions generated by the CID reaction leave the collision cell as one continuous ion beam. This beam is discretized into individual 'packs'. Each pack contains the entire set of product ions generated in the collision cell. The pack is accelerated orthogonally by an electric field from the pusher electrode. After a certain drift time, the flight direction of the ion pack is inverted through two consecutive electric fields opposite to the pusher electrode (two-stage ion mirror). The spectrum of flight times recorded at the detector can be translated into a spectrum of m/z values through calibration with a known mass standard.

Apart from full spectral acquisition, TOF mass detectors provide a much higher resolving power than quadrupole analyzers. Various definitions of resolving power exist, of which the nowadays more widely used definition is followed in this thesis. For a given isolated peak in a mass spectrum, its width

Δm is determined at half-maximal height (FWHM = full width at half maximum). The resolving power (RP_{FWHM}) is then defined as $RP_{FWHM} = \frac{m}{\Delta m}$, i.e. the quotient of the mass of the peak and its width at half-maximal height [67]. Commercial quadrupoles achieve a resolving power of 500-5000 (FWHM), whereas higher resolution comes only at the cost of decreased sensitivity such that operation at unit mass resolution is most appropriate [113]. On the other hand, TOF detectors nowadays achieve a RP_{FWHM} of more than 10,000. One factor limiting mass resolution in TOF MS is the initial spatial spread of the ion population in the acceleration region. Equation (2.10) includes this effect through the variable s , which represents the initial position of an ion in the acceleration field of strength E . The product of $E \cdot s$ is the effective voltage accelerating an ion. Any variance in initial position of ions of same mass, will translate into a variance in t_D . The second factor limiting mass resolution consists in the spread of initial velocities of the ions, i.e. the initial spread of kinetic energy which is represented by U_0 in equation (2.10) [67]. This energy adds to the kinetic energy imparted on the ions through the acceleration field and therefore also causes a temporal spread in arrival time at the detector. This distribution in arrival times translates into a distribution in the m/z domain, resulting in a mass peak with a full width at half maximum Δm .

$$t_D = \frac{D}{2} \sqrt{\frac{2 \cdot m}{U_0 + q \cdot E \cdot s}} \quad (2.10)$$

To best possibly focus ions of equal m/z but different initial kinetic energies, so called ion mirrors are employed (see Figure 2.6). An ion mirror consists of at least one but more often several electric fields which decelerate the ions and eventually invert their flight direction. During this process ions of higher initial kinetic energy penetrate the inversion field deeper than ions of lower initial kinetic energy but same m/z [48]. This process focuses ions of equal m/z value but different initial kinetic energies. The QqTOF MS device used in this thesis, the TripleTOF 6600 of AB Sciex, actually employs a two stage ion mirror based on two subsequent inversion fields of different voltages [8]. An optimal setting of these voltages allows to minimize second-order contributions of the initial kinetic energy distribution to the arrival time spread. Equipped with this sort of technology, the TripleTOF 6600 achieves a RP_{FWHM} of more than 30,000 for the m/z 397.2122 product ion of reserpine [3].

Resolving power as high as 30,000 enables one to resolve many analytes with equal nominal m/z value. This ability boosts the selectivity of a given analysis in complex matrices and increases accuracy by excluding the contribution of other ions to a given targeted mass transition. Most importantly, resolving power is a necessary though not sufficient prerequisite for accurate mass estimation. The accurate mass is estimated as a first moment (centroid) of the mass peak in question. The precision of this estimation is proportional to $\sqrt{h/\Delta m}$, where h is the peak height [56]. Hence, mass resolving power is necessary for accurate mass estimation. To actually obtain accurate masses, a proper calibration of the ion flight times needs to be performed.

5 LC-ESI-MS/MS-based quantitative Biology

5.1 LC-ESI-MS/MS-based Metabolomics

LC-ESI-MS/MS has been extensively used in targeted metabolomics, which represents the conceptually easiest application field in quantitative omics. Targeted metabolomics approaches rely on acquiring the signal on one or a few mass traces of a predefined set of metabolites [25]. The mass transitions for the metabolites of this set are identified from the acquisition of a reference spectrum upon injection of an analytical standard of the compound in question or a chemical similar one. From this spectrum usually the most intense ion is selected, for which the parameters of the ion optics are optimized for maximum sensitivity [13]. In LC-ESI-MS/MS-based metabolomics fragmentation is performed to increase specificity through the structure-dependent fragmentation and to improve the signal-to-noise ratio by the filtering operation of an additional quadrupole on the chemical background [174]. Therefore, with regard to the product ion spectrum the only properties of interest are signal intensity and the mass of product ions. The signal can be acquired without knowing the structural identity of the product ion as long as the fragment provides sufficient intensity.

The signal on a mass trace of one particular metabolite may serve to proof its presence. For accurate quantification, however, the raw mass trace intensity of a metabolite cannot generally be calibrated by a concentration standard. Due to non-constant ion source performance and ion suppression effects outlined in the Introduction, the use of internal standards (IS) for quantification purposes is considered mandatory [200]. With identical IS spiked into concentration standards and samples alike, the measured quantity of interest is the ratio of the fragment ion trace intensities corresponding to the analyte and the IS [200, 174]. Under the assumption that the fluctuation of ionization efficiency affects analyte and IS alike, the ratio of both mass trace intensities ought to be invariant under the above mentioned factors.

The assumption of equal ionization behavior of analyte and IS is justified as long as the internal standard is chemically comparable to the analyte in question. Deuterated analogs of a single amino acid or several amino acids may be used as IS [174]. However, the resulting mass shift of the IS through heavy hydrogen atoms is usually low and pronounced deuterium isotope effects with respect to the chromatographic separation may cause analyte and IS to elute at different times [42]. As a more attractive alternative, ^{13}C -labeled analogs of the metabolites to be quantified are not only chemically fully equivalent, but also show a sufficient mass shift and no altered chromatographic retention. Due to the high cost of authentic ^{13}C -labeled standards, the biotechnological production of them is often most appropriate. Wu et al. [200] used cell extracts of *Saccharomyces cerevisiae* grown on $[\text{U-}^{13}\text{C}]$ glucose as internal standards and obtained coefficients of determination of 0.99 for the calibration curves of central metabolism intermediates.

It is well established that microbial primary metabolism reacts in time scales of seconds to environmental perturbations. Therefore, rapid quenching of the metabolism in cold organic solvents is mandatory to obtain meaningful measurements, which are representative of the metabolism un-

der investigation [126]. However, during the quenching event some intracellular metabolites 'leak' into the quenching solvent. Thus, to obtain accurate estimates of intracellular concentrations, this effect needs to be corrected for. The use of quenching and internal standards notwithstanding, quantitative metabolomics suffers from relatively large errors [133]. These errors can firstly be attributed to metabolite instability [164] since many metabolites of the primary metabolism such as ribonucleoside triphosphates, reduced nicotinamide dinucleotide phosphates and α -keto acids are quite unstable. A further reason for the large errors on the finally calculated concentration values is due to the fact that concentration measurements in three samples are needed to account for 'leakage' effect [176].

5.2 LC-ESI-MS/MS-based Proteomics

Proteomics is defined as the identification and quantification of proteins and peptides. For many years now, gel-based approaches have increasingly been replaced by methods based on LC-ESI-MS/MS instrumentation. This trend has brought forth LC-ESI-MS/MS quantitative proteomics protocols for all well-studied model-organisms of biology [55, 183, 159]. The triumph of LC-ESI-MS/MS in the realm of proteomics is due to the quantification capabilities, sensitivity and specificity of LC-ESI-MS/MS-based protocols.

The specificity of LC-ESI-MS/MS workflows is mainly brought about by quantifying peptides derived by specific amide bond cleavages from the proteins of interest. The proteolysis is most commonly achieved by employing trypsin as endopeptidase. The cleavage specificity of this enzyme for amide bonds involving L-arginine (ARG) or LYS residues generates mostly peptides that are unique to the corresponding protein in a given proteomic background. Notwithstanding, peptides to be included for quantitative measurements need to be examined for uniqueness for the protein they are supposed to quantify. The proteolytic peptides are separated using reversed-phase HPLC with acidic mobile phase. Upon ionization the majority of peptides arise as doubly, triply and fourfold charged ions from the ESI process, one of which is usually more intense than others [124]. Figure 2.7 A depicts the charge state +2 for the tryptic peptide GISLVPAFSAAER derived from fatty acid synthase (cg0957) from *C. glutamicum* for the purpose of illustration.

In contrast to the fragmentation reactions of metabolites, the rules governing the CID reaction of peptide ions are rather straightforward [70]. The most frequently observed and intense fragment ions arise from the cleavage of the amide bond of the amino acid backbone which is exemplarily depicted in Figure 2.7 B. When the charge is retained on the amino acid of the C-terminal end (right), y-ions are formed (see Figure 2.7 B, whereas b-ions are formed when the charge migrates to the oxygen atom of the amino acid of the N-terminus (see Figure 2.7 C). Figure 2.7 C also shows that when several charges are present on the peptide, both ions can be generated from one fragmentation. Both y- and b-ions can be further classified according to the number of amino acid residues they retain and the charge they bear. Upon CID fragmentation of peptides, a series of b- and y-ions is formed which is specific for the sequence of the fragmented peptide.

Given this specificity inherent in the sequence-specific tryptic digestion and CID fragmentation, MRM protocols for quantitative targeted proteomics are based on selecting a set of peptides of the proteins of interest and a corresponding set of fragment ions for each peptide. The peptides to be included can be retrieved from public repositories [147]. Alternatively, a prior information-dependent acquisition of the sample identifies intense peptide and fragment ions [76]. To afford sufficient specificity, it is generally recommended to monitor at least three transitions per peptide [110]. Given the defined set of peptides and fragments the arising mass transitions are defined in an MRM assay. The set of transitions may also contain transitions corresponding to labeled isotopes of the chosen peptides, depending on the internal standardization (see below). Due to the large number of mass transitions that need to be measured to ensure selectivity, MRM assays are limited with respect to the number of quantifiable peptides. The cycle time limit outlined in Section 4.2 limits the number of measured peptides in one acquisition to several hundreds at most [161]. Therefore, MRM protocols are usually restricted to a small part of the metabolism [183] or certain signal transduction pathways.

Surely, QqTOF MS devices may be employed to circumvent the cycle time limitation with respect to product ion detection for each Q1-selected precursor peptide as outlined in Section 4.3. However, when precursor ion selection is still performed with unit resolution, the cycle time limit imposed by the first quadrupole for precursor ion selection would remain in place, limiting the overall performance in fast acquisition of peptide spectra. To fully circumvent the Q1 cycle time limit, a pioneering approach has emerged: Sequential window acquisition of all theoretical fragment ion spectra (SWATH) [61] (Figure 2.8). In this setting the resolution of the Q1 quadrupole is deliberately worsened from unit resolution conventionally employed in MRM assays to allow a mass range of 10-25 Da to pass [76] such that several precursor ions may be co-isolated at any given point in time (see Figure 2.8 A). Figure 2.8 B) shows the case where the isolation window width is constant over the m/z domain. Each such window is called *swath* [61] and is depicted as black bar along the m/z domain in Figure 2.8 B). Using such an isolation window width, the entire m/z range in which most peptides lie (m/z 400-1200) can be cycled through in 3.2 seconds, well below the chromatographic elution time of peptides [61, 161] and several orders of magnitude faster in comparison to unit resolution. However, since several precursor ion species are likely to be co-isolated in each swath (see Figure 2.8 A), the acquired TOF product ion spectrum of each swath is a superposition of several product ion spectra and consequently significantly less specific in comparison to the scenario of unit resolution precursor isolation.

Although neither the quadrupole Q1 nor the TOF MS isolate a single precursor and fragment ion pair, the concept of a mass trace is still applicable. The mass of each fragment ion of the targeted peptides is extracted from the MS2 spectra. The probability of co-extracting an interfering mass is significantly reduced through the resolving power of the TOF detector. Its mass resolving power in the range of 30,000-35,000 is sufficient to compensate for the lower selectivity of the MS1 stage of mass analysis since mass peaks can be extracted using an extraction window as narrow as 10 ppm around the exact mass of the searched fragment ions [61]. When the corresponding masses are ex-

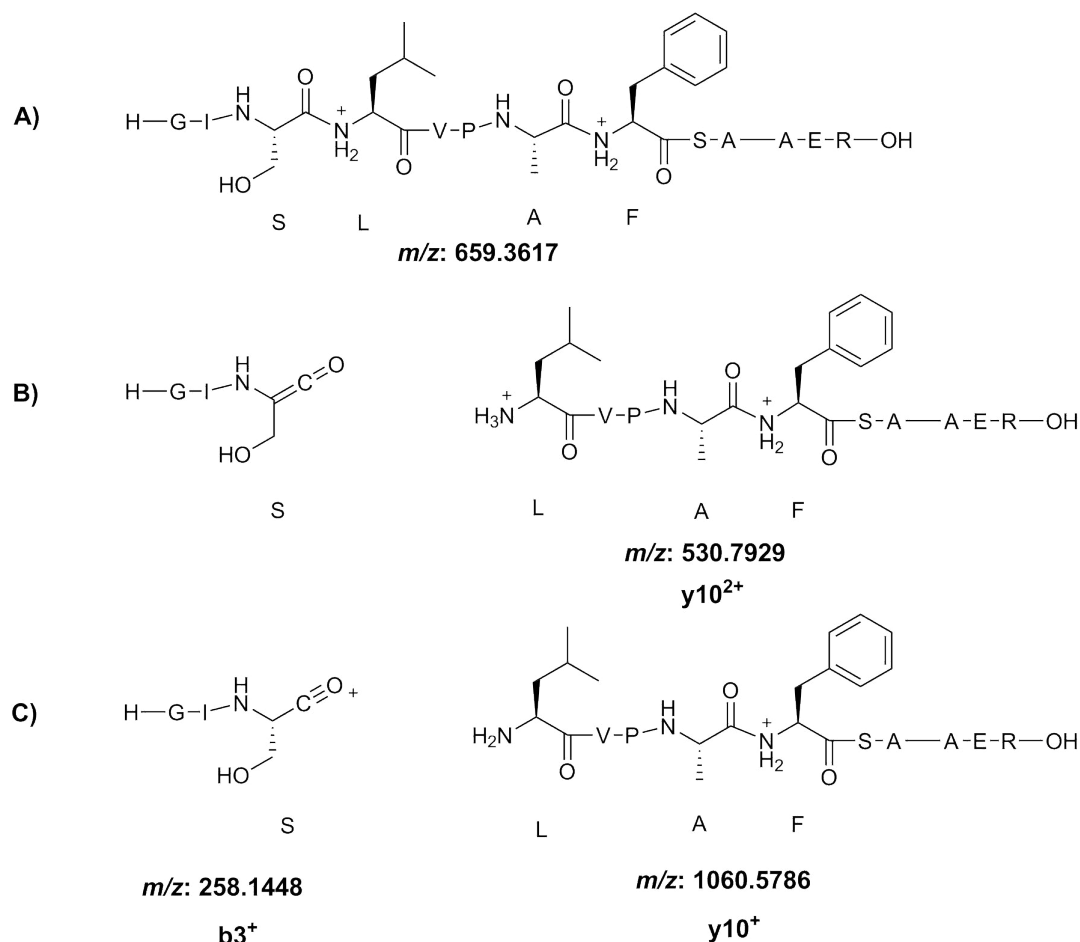


Figure 2.7: An exemplary CID fragmentation reaction of the doubly charged peptide GISLVPAFSAAER of protein cg0957 of *C. glutamicum*. A) The peptide is depicted with amino acids SER, LEU, ALA and PHE expanded as structures to indicate the position of the charge. B) When the charge is retained on the amino group of the amino acid from the C-terminus upon fragmentation, an y-ion is formed. Based on the number of retained amino acids and charge, it is denoted $y10^{2+}$. C) When the charge migrates to the oxygen atom of the N-terminal amino acid, a b-ion is formed.

tracted from sequential SWATH TOF spectra, the interference-free fragment ion chromatograms can be reconstructed post-acquisition [76]. This process is exemplarily depicted in Figure 2.8 B). The mass traces corresponding to the green peptide species can be reconstructed without interference, although the TOF spectra also contains fragment ion species from the red peptide ion. Using this approach, Gillet et al. [61] showed that an astounding 89 % of all peptides from a yeast proteomic background can be quantified with at least five interference-free transitions using 25 Da-wide Q1 isolation windows. Since one protein can be represented by various peptides, this technique provides almost genome-wide proteome coverage.

Even when a TOF extraction width as narrow as 10 ppm is used, it cannot be guaranteed that no other signal of fragment ions of co-isolated peptides contaminate the extracted signal of the targeted fragment. Therefore, all extracted fragment ion traces of one peptide are assembled into one so called *peak group* for which the several quality metrics are derived. Among others, peak similarity and co-

elution between fragment ion traces and match of measured accurate mass of the fragment ions are used to compute a score to judge the probability that the peptide was correctly identified by its fragment ion traces [61]. The integrated ion trace intensities, on the other hand, quantify the peptide. Therefore, quantification and identification are performed with almost genome-wide coverage.

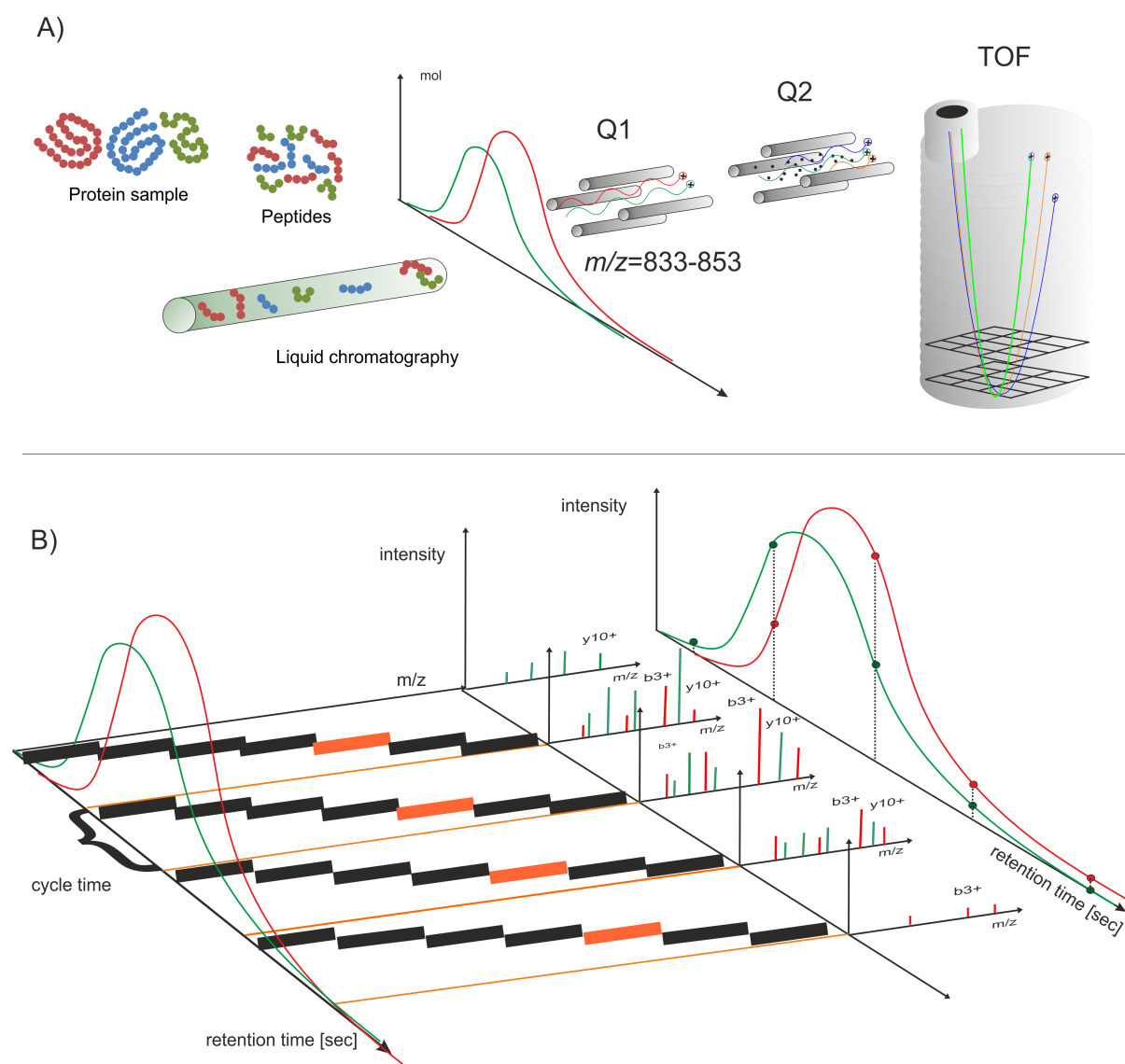


Figure 2.8: QqTOF MS-based proteomics workflow employing SWATH MS acquisition. A) The eluting peptides are mass-filtered in the first quadrupole. Since the transmission window is several Da wide, a multitude of peptide ions are co-isolated and transmitted concurrently to the collision cell and fragmented. The TOF detector records fragments from all co-isolated precursor ion species. B) The scanned Q1 window from subfigure A) is just one of several so-called *swaths* which are consecutively scanned over the entire m/z domain of interest. Each swath is represented as black bar. After one cycle time the sequence of swaths is scanned in a new cycle. For the swath designated with an orange bar the convoluted spectrum is shown to the right of subfigure B. From the spectra of the same swath acquired over time, specific fragment ions corresponding to a desired peptide can be extracted in a targeted manner. Due to the high resolution afforded by the TOF detector, many fragment ion traces can be extracted without interferences.

Regardless of the MS acquisition method (be it MRM or SWATH) an appropriate abundance feature has to be selected when it comes to quantification. As outlined above, the ion count of fragments is used as abundance feature in SWATH MS. The same holds true for MRM assays. Methods based on ion intensities are amenable to the use of internal isotopic standards. Broadly, two categories have emerged for internal standardization in proteomics: metabolic labeling and chemical labeling.

Chemical labeling strategies use distinct isotopomers of chemical derivatization reagents to introduce a defined mass shift to the proteome sample of one condition [58]. The isotope label may be introduced at the level of proteins [162] or peptides [72] of a sample through a chemical reaction. Since internal standards can only account for experimental variation downstream to their introduction into the sample processing, peptide chemical labeling approaches cannot account for systematic effects occurring during trypsinization of intact proteins. Samples representing different conditions to be compared are derivatized with different isotopomers of the same chemical derivatization agent [27]. On the MS level each condition then corresponds to the signal on a distinct mass trace of a fragment ion retaining the chemical label. Thus, the number of different isotopomers available for a given reagent determines the 'plexes', the number of distinct conditions distinguishable in one sample. The most popular chemical labeling reagent class are iTRAQ reagents which nowadays offer eight plexes [72]. Since each isotopomer of the derivatized proteins or peptides is subject to the same ion suppression effect and instrument variation, the relative intensities of the mass traces corresponding to the employed isotopomers yield an unbiased estimate of relative expression between the investigated conditions. Since usually several peptides and fragments are available for one protein, the isotope ratio can be determined from the corresponding mass traces of several fragments. Since each fragment or peptide optimally reflects the same isotope ratio, averaging over all measured fragments or peptides can be performed.

On the other hand, metabolic labeling introduces the labeling at the level of proteins at the cultivation level prior to any sample processing. Essentially, two methods have emerged: labeling with [U-¹⁵N-¹³C]LYS and [U-¹⁵N-¹³C]ARG in organisms/cells auxotrophic for both amino acids [77] and ¹⁵N-labeling for non-amino acid auxotrophes, notably prokaryotes [27]. In each of the techniques only two labeling states are generated per peptide, i.e. light and heavy. Therefore, only two plexes per sample are available. This has been noted as particular disadvantage of this technique. A particular advantage, however, consists in introducing the internal standard at the earliest possible time point. Heavy and light isotopes can be mixed before trypsinization and therefore account for all possible artifacts during proteolytic digest. Furthermore, this technique does not require expensive chemical reagents and no further steps beyond the ordinary sample preparation.

Given the chosen signal and labeling strategy, a diverse set of data processing approaches exist to compute relative changes and associated statistics. From the typical LC-ESI-MS/MS proteomics workflow in Figure 2.8 it becomes apparent that several intensity features are observed for one protein. The analytes in this LC-ESI-MS/MS analysis are peptides, the abundance of which is quantified from their CID fragments. The quantity of biological interest, however, is the amount of the corre-

sponding protein, which each workflow ultimately ought to quantify. Therefore, a rational approach to data integration is needed. A sensible approach would be to sum all signal intensity over all the fragments of all peptides from one protein. Although this approach may be workable for label-free quantitative proteomics, it is not very useful for quantification involving internal standards.

Internal standards serve to compensate for ion suppression effects that affect the analyte (peptide) intensity such that the light-to-heavy (L/H) ratios need to be computed at least on the peptide level. Further down the hierarchy, L/H ratios can be computed on the level of each mass transition. No matter on which level the L/H ratios are computed, they need to be combined appropriately to arrive at protein level conclusions. The averaging can be done in a weighted manner, incorporating the signal-to-noise ratio of the integrated peak and other sources of uncertainty [86]. However, these procedures are heuristic in nature and have no foundation in statistical theory. Moreover, the standard parametric statistical tests cannot be applied to the L/H ratios: If the peak areas are assumed to follow a normal distribution, then the ratio of both areas follows a Cauchy-distribution.

Alternatively, regression models reconcile all data in one estimation. These models take into account each factor affecting the measured fragment ion trace intensity. Under the assumption that each factor acts multiplicatively on the observed peak areas, an additive model arises after log-transformation of the raw peak areas. Thus, the log-transformed peak area can be modeled as linear additive regression model [32, 92]. The most basic form of this regression problem models peak areas acquired on peptides from LC-ESI-MS according to equation (2.11) [92]:

$$y_{z,p,t} = \tau + \Pi_p + T_t + \epsilon_{z,p,t} \quad (2.11)$$

τ is the overall mean of the vector $\mathbf{y}$ of all data points for the considered protein, Π_p is the effect of the p^{th} peptide and T_t the effect of the t^{th} treatment. Published approaches differ in the extent to which different factors are included into the model. ^{14}N areas can be modeled as linear additive model in the software package MSstats [32]. For isotope ratio estimation, on the other hand, Hill et al. [74] applied a regression approach to model isotope ratios in samples which were derivatized with the iTRAQ reagents. In this framework, the isotope ratios directly represent the fold changes (FCs) between the conditions to be compared.

5.3 ^{13}C -Fluxomics

5.3.1 Overview on ^{13}C -Metabolic Flux Analysis

In contradistinction to MS-based acquisition methods for metabolomics and proteomics, which directly measure the quantity of interest or a quantity proportional to it, intracellular metabolic fluxes cannot be directly assessed *in vivo*. In fact, fluxes can only be inferred from the integration of quantitative measurements into suitable biochemical network models. In particular, ^{13}C -MFA determines intracellular fluxes from measurements of extracellular rates and ^{13}C -label enrichment in intracellular metabolites fitted to a carbon atom transition model [99, 28, 203, 188]. The latter are generated

in a cultivation experiment employing selectively ^{13}C -labeled growth substrates and are most commonly measured by MS-based techniques. The fluxes are then inferred from these data based on a mathematical network model describing the distribution of isotopes over a metabolic network in dependence of the prevailing fluxes. The labeling experiment in combination with the model-based data evaluation is known as ^{13}C -MFA [132]. Figure 2.9 depicts the workflow of a full-fledged and comprehensive ^{13}C -MFA study. Due to the complexity of ^{13}C -MFA as method, several of its six pillars represent proper fields of inquiry, with most research devoted to pillar D and E. Therefore, the following subsection is devoted to a detailed description of each of the six pillars.

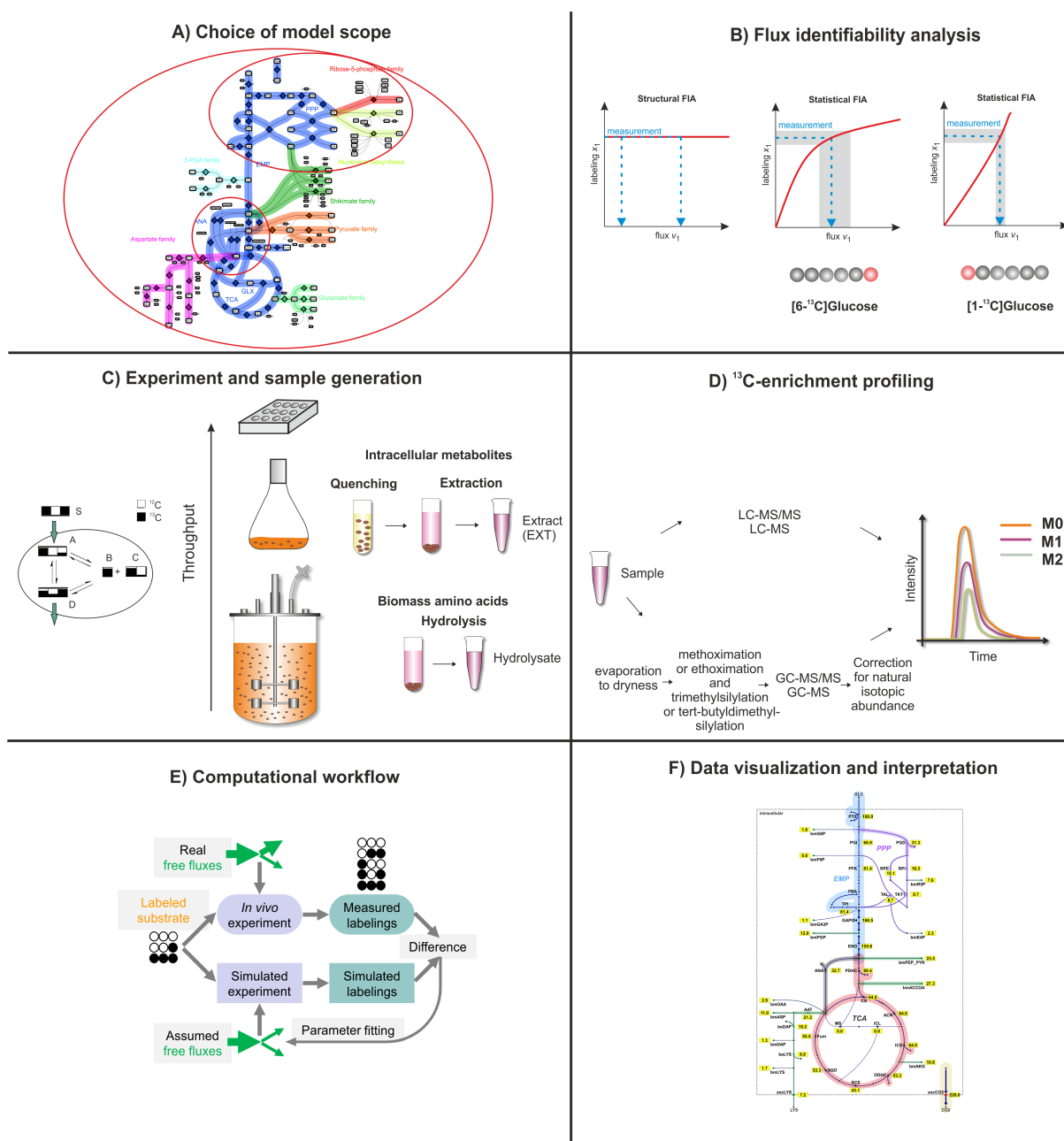


Figure 2.9: The workflow of a full-fledged ^{13}C -MFA study.

5.3.2 Choice of model scope

Every ^{13}C -MFA study is performed with a research question in mind. The focus may be the central carbon metabolism as a whole [179] or a specific pathway [69, 16]. Depending on the scope of the research question, the model size and the setting of system boundaries have to be chosen accordingly (represented by the differently sized ellipses in Figure 2.9 A). Regardless of the scope, the carbon atom transitions for all reactions of the considered subsection of metabolism need to be known (see below). This requirement has ^{13}C -MFA historically restrained to the central carbon metabolism since the carbon atom transitions apart from the central pathways have not been worked out accurately. However, recently attempts were made at expanding ^{13}C -MFA beyond the central carbon metabolism to include many more biomass and cofactor synthesis reactions [120].

5.3.3 Flux identifiability analysis

With the model size defined, one particular but non-unique set of free fluxes must be chosen. The free fluxes represent the missing flux information and may be inferred from fitting an isotopomer model of the metabolic network in question to ^{13}C -labeling measurements. Given the ^{13}C -label measurements that are analytically accessible in the considered section of metabolism, one has to ascertain *a priori* whether the fluxes will be identifiable. This analysis is called structural *Flux Identifiability Analysis* (FIA). This step is seldom done *a priori* under the assumption that the intracellular flux distributions can be inferred if sufficiently informative labeling measurements are provided. The majority of ^{13}C -MFA studies focus on an *a posteriori* statistical identifiability analysis [179, 134, 38], in which confidence intervals for flux estimators are derived based on real measurement data. Two of these studies obtained *a posteriori* non-identifiable fluxes for anaplerotic reactions in *C. glutamicum* based on standard deviations several times greater than the flux value. Inflated confidence intervals may be due to a low sensitivity of the flux variables to observable measurements (statistical identifiability) or may alternatively be a numerical reflection of structural (i.e. model-inherent) non-identifiability by ^{13}C -MFA. If the latter is the case, either structurally more informative measurements need to be performed or the model must be reduced. If the measurement of the total isotopomer vector of each metabolite does not yield structurally identifiable fluxes, no choice of tracer or measurement setup will allow to infer the affected fluxes. Thus, many unnecessary experiments can be avoided by a simple *a priori* analysis.

Approaches to *a priori* identifiability problems may be categorized according to properties of the model under investigation or properties of the method itself. To distinguish these approaches the following terminology will be henceforth used:

- *Structural FIA* is the most fundamental one because it analyzes if there is a problem in the model itself (Figure 2.10 A). To this end, FIA starts with the best case assumption that no measurement errors are present (noise-free data). Noise-free data are generated from the model in question by means of a forward simulation. A structural analysis then investigates if the pa-

rameters of interest (fluxes) can be calculated from the model equations given the exact and noise-free data. For classical ^{13}C -MFA the model equations are isotopomer balance equations under metabolic steady state. However, even if fluxes are structurally identifiable reasonable confidence intervals may not be obtained in case of realistic measurement noise. This question lies at the center of *statistical* FIA (Figure 2.10 B). Of course, if fluxes are structurally non-identifiable then they are also not statistically identifiable, which is why a structural FIA has to precede every sound statistical FIA. An application of statistical FIA is Experimental Design for optimal tracer choice [127, 9] which, however, is not subject of this thesis.

- Answering identifiability of fluxes at one fixed flux and/or input configuration amounts to performing a *local* FIA the result of which is only valid for the particular point of operation. It can therefore only prove that no other flux solution in the direct neighborhood of the calculated exists. Whether local identifiability holds over the entire feasible domain of fluxes and/or input configurations is answered by *global* FIA (Figure 2.10 C).
- Even if a single locally well determined flux solution can be found at every point of the flux space, it is not yet clear whether the obtained flux is uniquely determined over its entire domain by the particular labeling measurement. Therefore, *unique* FIA tries to prove that exactly one single set of intracellular fluxes exists which produces a given set of measurements (Figure 2.10 D). Unfortunately, this ambitious goal can only be reached in some rare cases. Particularly, all methods based on Jacobian matrices are *non-unique* FIA methods because they rely on a linear approximation of the investigated model around a point in parameter space. The non-existence of other solutions can only be checked empirically.
- *Full information* FIA assumes an ideal measurement configuration in which every intracellular labeling enrichment is observable. Clearly, this is never the case in practice. Depending on the analytical platform (i.e. LC-ESI-MS/MS, GC-EI-MS, GC-EI-MS/MS or NMR) the label enrichment in different parts of the carbon backbone may be experimentally accessible. Even if the entire label information of certain metabolites can be obtained, as demonstrated for ASP [29], at most the same flux information compared to a full information analysis can be obtained. Thus, if fluxes are non-identifiable with full information, then this also holds in the case of partial information in any possible real measurement configuration. The latter case is the topic of *partial information* FIA.
- All FIA methods applied in the following rely on matrix rank computations which is a notoriously error-prone numerical procedure [54]. This means that flux identifiability based on *numerical* FIA using standard floating-point arithmetics can never be doubtfully decided if a matrix is singular or just close to singular. This might be the result of a poor experimental design but not a structural problem. As a powerful but computationally expensive and memory consuming tool *exact* FIA based on rational number arithmetics can rigorously determine the

rank of a matrix and, thus, is an invaluable tool in structural FIA. However, performing exact integer computations on a large ^{13}C -labeling network is extremely memory consuming. Thus, in practice this kind of analysis is restricted to smaller networks. Furthermore, any ordinary differential equation system cannot be analyzed with integer arithmetics due to the need for numerical integration (i.e. approximation methods).

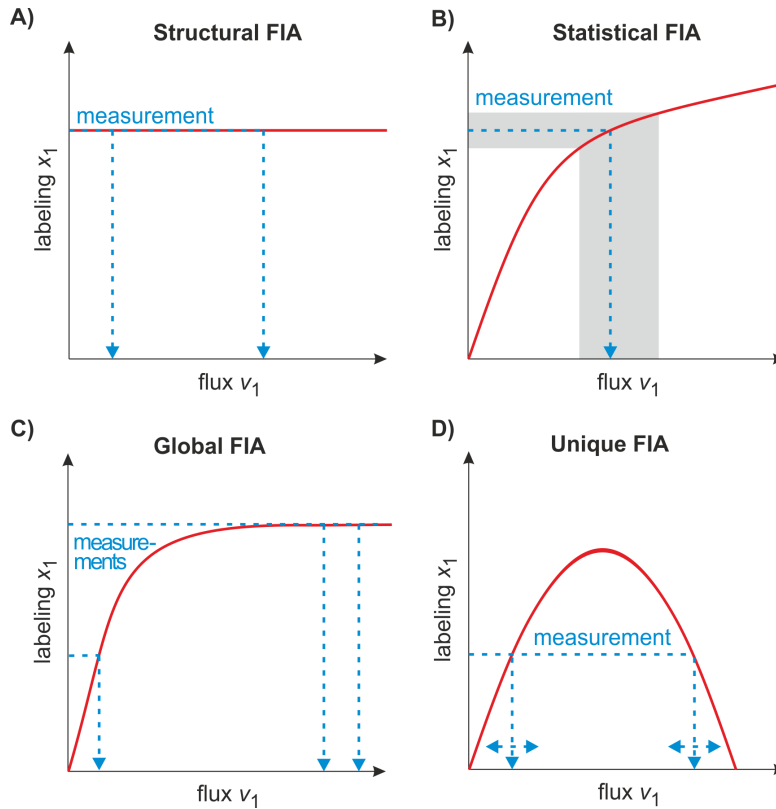


Figure 2.10: Different identifiability concepts: A) Structural FIA: The flux is not determined by every possible measurement. B) Statistical FIA: Even if the flux is structurally determined by the measurement, measurement errors lead to a poor statistical quality (shaded area). C) Global FIA: Local Invertibility of the model is given only on a limited region in the flux space. For some flux values the model does not allow an inversion (horizontal slope on the r.h.s.). D) Unique FIA: Even if the model is locally invertible, the obtained flux for a given measurement is non-unique. Only in the neighborhood of a selected working point is the solution unique (double arrows).

In any case the result of a FIA is expressed by the characteristic number $rDOF$, the remaining degrees of freedom. $rDOF$ essentially gives the number of intracellular fluxes which cannot be determined by the experiment. More precisely, there exists a subset of $rDOF$ intracellular fluxes for which the values must be given from additional information not contained in the measurements in order to obtain a single solution. Thus, flux identifiability is given if $rDOF = 0$ and any higher number indicates how far away we are from complete identifiability.

To compute the $rDOF$, one has to determine the rank of a method-dependent matrix. If this matrix contains only rational numbers, its rank can be exactly determined using rational arithmetics. Oth-

erwise, the most stable and precise numerical method for rank estimation is the singular value decomposition (SVD). SVD yields a spectrum of positive singular values, ordered by magnitude, and the number of zero singular values indicates the matrix rank deficiency. In practice a computed singular value is never exactly zero due to the limited precision of floating point arithmetics. Thus, a threshold must be chosen to distinguish those singular values 'close to zero' from the significant ones. However, a clear separation is not always possible.

5.3.4 ^{13}C -labeling experiment

The equations on which ^{13}C -MFA relies relate flux values to relative isotopomer abundances. This isotopomer distribution needs to be generated experimentally during a cultivation of the organism in question on ^{13}C -labeled substrates. The employed cultivation conditions need to ensure, regardless of the cultivation platform, that the basic working assumption of ^{13}C -MFA is met: stationary metabolite pools (see below). Although metabolic stationary conditions are in theory only attainable in chemostat cultivation mode, the majority of ^{13}C -studies recur to batch cultivation [101, 73, 179, 136, 120, 115, 11, 34]. Under this condition the metabolism is assumed to be in *quasi-stationary* state, i.e. an exponentially growing culture with constant biomass specific rates of substrate consumption and product formation.

For ^{13}C -labeling experiments in batch mode shaking flasks are often employed due to their simplicity and easy setup [136, 120, 115]. However, to achieve best control over the experimental conditions, bioreactor cultivations are more appropriate, allowing the measurement and control of pH and dissolved oxygen (DO) [179]. Furthermore, only closed bioreactor cultivation systems with controlled aeration and exhaust gas analysis allow to accurately balance inflows and outflows of all carbon atoms. This balancing step is almost never performed, but crucial because it allows to check experimentally determined extracellular rates for consistency [100]. A closed carbon balance is the best indicator that the proposed model of metabolism accounts for all major (by)product formation reactions [24].

Since bench-top bioreactor cultivations entail significant material and time effort, smaller setups with increased throughput were realized through miniaturization by several research groups [108, 100]. Most prominently amongst them, the Antoniewicz lab employs Hungate tubes with 10-15 ml working volume with controlled aeration using high precision pumps [34]. Most recently, the miniaturization has been driven to the μl -scale [101, 73] in microtiterplates (MTP). Current MTP cultivation systems allow for optical continuous monitoring of microbial growth, pH and DO using optodes. To some extent, these systems also allow basic control over some process parameters although feedback control is mostly not supported. When integrated into robotic liquid handling systems, these platforms allow unprecedented throughput by automating sample generation and processing [73] (see ranking of cultivation systems in Figure 2.9 C).

5.3.5 ^{13}C -isotopomer profiling

Regardless of the scale, each experiment needs to yield samples the ^{13}C -distribution of which is representative of the flux distribution prevailing under the employed cultivation conditions. When ^{13}C -label is to be measured directly in primary metabolites from intracellular extracts, microbial cultures need to be quenched rapidly [126]. A prime reason for this requirement is that primary prokaryotic metabolism responds in sub-minute time scales to environmental perturbations inherent in sample processing [135, 200, 125]. Intracellular extracts allow to profile ^{13}C -incorporation in metabolites that are not precursor metabolites to amino acid biosynthesis and consequently have no corresponding analog in amino acids such as MAL and GLX, CIT, ICIT, cis-ACO, fructose 1,6-bisphosphate (FBP) and sedoheptulose 7-phosphate (S7P). Furthermore, the time-resolved profiling of ^{13}C -incorporation in primary metabolites is a prerequisite for isotopically instationary experiments [135]. For stationary ^{13}C -MFA, on the other hand, ^{13}C -profiling is most commonly performed in amino acids from hydrolyzed biomass protein [196, 206], which does not entail the intricacies of quenching of intracellular extracts. Both methods yield a sample for analysis (see Figure 2.9 D).

The profiling of ^{13}C -label distribution in samples from both approaches is most commonly accomplished by MS. However, prior to MS detection sample matrices need to be chromatographically separated to reduce spectral complexity and in the case of LC-ESI-systems to avoid ion suppression. When it comes to chromatographic separation ^{13}C -MFA studies recur to either gas chromatography (GC) or LC.

For GC-MS detection, metabolites from the primary metabolism need to be derivatized to render them volatile enough for GC separation. This can be achieved by converting them to either by trimethylsilyl (TMS) [115] or *tert*-butyldimethylsilyl (TBDMS) derivatives [206]. The amino, carboxyl, phosphate, sulfhydryl and hydroxyl groups of these molecules are derivatized with either $\text{Si}(\text{CH}_3)_3$ or $\text{Si}(\text{CH}_3)_2-\text{C}(\text{CH}_3)_3$, respectively. Consequently, this derivatization is sufficient for all amino acids and hydroxy carboxylic acids. However, when sugars or keto-acids are to be measured, additional derivatization steps need to be performed before silylation can be carried out such as oximation [106], methoximation [202] or ethoximation [115]. The molecules arising from these derivatization steps exhibit a significant mass shift. Moreover, the ^{13}C -distribution in the carbon backbone of the metabolites is convoluted with the natural isotopic distribution of the carbon and silicon atoms of the derivatization agents, such that the fractional abundance (FA) of mass traces does not directly reflect mass isotopomer abundance. The ^{13}C -distribution of the carbon backbone needs to be restored through a correction process [87].

Though GC-MS technology slightly prevails in frequency of application for ^{13}C -profiling [206], LC-MS technology is widely used to quantify ^{13}C -label enrichment in central metabolism intermediates [125, 79, 95, 179, 156, 120, 7] and will continue to be used due to some inherent advantages:

- Cell extracts do not need to be evaporated to dryness in order to perform methoximation or ethoximation and subsequent TMS or TBDMS [82, 202, 115]. Although the derivatization steps

may be automated [30], the total sample processing time in LC-ESI-MS/MS is inherently shorter since cell extracts can be injected directly, at the cost of retention time reproducibility (see Figure 2.9 D).

- LC enables to separate a large part of the central carbon metabolism intermediates in a single chromatographic run using hydrophilic interaction chromatography [60, 170, 205].
- Correcting the obtained isotopomer distribution from LC-ESI-MS/MS or LC-ESI-MS for the occurrence of natural isotopes is inherently more precise and less biased due to the absence of heterogroups from derivatization agents. The correction process of GC-MS/MS mass fractions was shown to inflate the coefficient of variation for low intensity (tandem) mass isotopomer fractions [115].
- Whereas LC-MS is only slightly more popular than GC-MS in the realm of metabolomics [26], its lead in quantitative proteomics is unrivaled [159, 55]. As such, an LC-ESI-MS/MS device can serve as one-stop platform for proteomics, metabolomics as well as fluxomics, requiring just a change of column between proteomics applications on the one hand and metabolomics and fluxomics experiments on the other hand.

After chromatographic separation, subsequent mass analysis allows to determine relative abundances of ^{13}C -labeled species. From the total labeling state of a metabolite, represented by the vector $\mathbf{x}$ containing its 2^n isotopomers ordered by binary value, usually only a few isotopomers of a metabolite can be analytically disentangled. In general, each measurement setup, be it GC-MS, GC/LC-MS/MS or NMR [23] can be described by a function $\mathbf{g}(\mathbf{x})$, mapping the isotopomer vector $\mathbf{x}$ onto the label measurements $\mathbf{y}$. For MS or MS/MS data, in particular, the function $\mathbf{y} = \mathbf{g}(\mathbf{x})$ takes the form of a linear mapping from $\mathbf{x}$ to $\mathbf{y}$ according to:

$$\mathbf{y} = \mathbf{C} \cdot \mathbf{x} \quad (2.12)$$

The aim of any ^{13}C -profiling technique is to gain as much information about the labeling state $\mathbf{x}$ of a metabolite as possible. The information content of $\mathbf{y}$ with respect to $\mathbf{x}$ is captured by the rank of matrix $\mathbf{C}$. Established ^{13}C -profiling techniques differ in the rank of matrix $\mathbf{C}$. The most widely employed method for ^{13}C -isotopomer profiling is GC-EI-MS of TBDMS-derivatized metabolites [41, 136, 202, 196, 206, 101, 34]. Upon EI metabolites usually fragment in the ionization source, providing rich product ion spectra [10]. This in-source fragmentation distributes the ionized amount of one analyte across several ions. For most TBDMS-amino acids the most intense fragment ion corresponds to the loss of a *tert*-butyl-radical, yielding an ion at $[\text{M}-57]^+$ [10, 136]. This ion retains the full carbon backbone. When ^{13}C -enrichment is measured in this ion, just the molecules having the same number of ^{13}C -labeled atoms can be distinguished. These entities are called mass isotopomers (or isotopologues). The resulting vector of relative intensities is called Mass Isotopomer Distribution

(MID). This distribution does not provide any information about the position of labeled carbon atoms within the precursor ion.

Many EI-MS spectra of TBDMS-derivatized metabolites exhibit in-source fragments retaining just a part of the carbon skeleton. ^{13}C -profiling in these fragments is less informative since fewer ^{13}C -labeling states are possible in the fragment ion than in the intact precursor metabolite. However, this has been circumvented by including the ^{13}C -label distribution of fragment ions from other carbon backbone breaks. By combining the ^{13}C -enrichment from several such ions not only the limited information of each single ion can be augmented, but also positional information on the ^{13}C -label can be obtained [10, 41, 206]. Thus, through in-source fragmentation EI-GC may provide positional resolution in single-stage MS analysis. For some metabolites, however, ions with limited carbon backbone information are the only intense ions in their respective EI spectra. For instance, the sole intense ion of TYR-(TBDMS)₃ retains carbon atoms [1-2] [10], admitting three ^{13}C -labeling states. The rank of C for the measurement of ^{13}C -label in this fragment is three, whereas TYR has 2⁹ isotopomers. This is the only accessible information on the labeling state of TYR since no other intense EI-ions are generated for the TYR-(TBDMS)₃ derivative, which might contain more carbon backbone atoms. EI of sugar phosphates mostly yields ions with barely any information of the carbon skeleton [30, 115], limiting the applicability of GC-EI for glycolytic intermediates.

The ESI process in an LC-ESI-MS/MS analysis mostly preserves the precursor ion. Thus, through single stage mass analysis on these ions, only the MID is experimentally accessible. To obtain fragmentation, CID of intact precursor ions needs to be performed. The arising vector of ^{13}C -enrichment in one of the CID fragments, is referred to as tandem mass isotopomer distribution (TMID) [171]. In the same manner as in proteomics and metabolomics, MS/MS acquisition in ^{13}C -profiling confers specificity and increases sensitivity through reduction in chemical background. Analogously to GC-EI-MS analysis on fragments, positional resolution may be achieved when ^{13}C -distribution is measured in fragments from a carbon backbone break. In contradistinction to GC-EI, however, the MID is always contained in the TMID since the intact mass isotopomers are separately fragmented. This is graphically illustrated in Figure 2.11 for the m/z 88 fragment of ASP from Figure 2.5, stemming from carbon atoms [2-4]. In each case, the position of the ^{13}C in each MI is further resolved through the fragmentation. Additionally, all mass isotopomers are still resolved, contrary to GC-EI-MS. It may be noted that this scheme can only be derived by the knowledge of the positional origin of the carbon atoms retained in the fragment ion. Figure 2.11 also illustrates the notation for mass traces that will be henceforth used which is adapted from Choi and Antoniewicz [28]. A capital M denotes the MI with the number of labeled carbon atoms, whereas a lowercase m denotes a tandem mass isotopomer followed by the number of labeled carbon atoms.

The contribution of isotopomers to tandem mass isotopomers from Figure 2.11 can be coded as matrix mapping the isotopomer vector of a metabolite onto the vector representing the tandem mass isotopomer distribution:

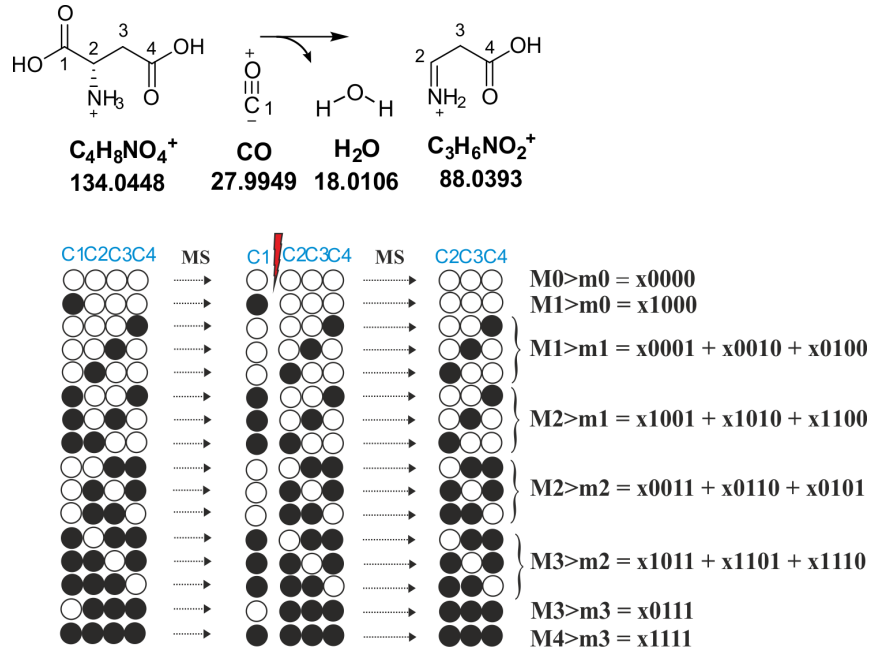


Figure 2.11: Relation between isotopomers of ASP and tandem mass isotopomers of the m/z 88 fragment of the $[M+H]^+$ ion of ASP. Please note that the isotopomers in this figure are ordered according to their contribution to tandem mass isotopomers. The tandem mass isotopomers, in turn, are ordered by the number of labeled carbon atoms from top to bottom. Thus, the ordering of isotopomers of ASP in this figure is different from the ordering in equation (2.13).

$$\begin{pmatrix}
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix} \cdot \begin{pmatrix} x_{0000} \\ x_{0001} \\ x_{0010} \\ x_{0011} \\ x_{0100} \\ x_{0101} \\ x_{0110} \\ x_{0111} \\ x_{1000} \\ x_{1001} \\ x_{1010} \\ x_{1011} \\ x_{1100} \\ x_{1101} \\ x_{1110} \\ x_{1111} \end{pmatrix} = \begin{pmatrix} M0 > m0 \\ M1 > m0 \\ M1 > m1 \\ M2 > m1 \\ M2 > m2 \\ M3 > m2 \\ M3 > m3 \\ M4 > m3 \end{pmatrix} \quad (2.13)$$

where x_{0000} , x_{0001} , etc. denote the individual isotopomers of ASP. From left to right, the positions of the subscript index denotes the carbon atom positions from C1 to C4 as in Figure 2.11. The entire

vector will be denoted $\mathbf{x}_{\text{ASP}}$. From equation (2.13) it becomes apparent that the isotopomer x_{0111} is directly resolved from all other isotopomers as is x_{1000} . By measuring the ^{13}C -enrichment in a fragment retaining all carbon atoms, both would be aliased with all other isotopomers bearing the same number of labeled carbon atoms.

The improved isotopomer resolution afforded by MS/MS in LC-ESI-MS/MS or GC-EI-MS from several fragments with part of the carbon backbone appears to translate itself to an improved statistical quality of flux estimators. This is the reason as to why ever greater efforts have been made in recent years at measuring ^{13}C -enrichment in fragments originating from a carbon-backbone cleavage [156]. The advantage of positionally resolved ^{13}C -label information for increased precision of flux estimation was shown experimentally for both GC-EI-MS/MS and LC-ESI-MS/MS [83, 156, 136]. Recent simulation studies demonstrated the gain in statistical quality of flux estimators using positional ^{13}C -enrichment of molecules with more than three carbon atoms [28] and even for situations in which the carbon origin is not known [171]. Moreover, some pathway fluxes (like Entner-Doudoroff and Embden-Meyerhof-Parnas) can only structurally be identified from ^{13}C -labeling data using positionally resolved ^{13}C -enrichment in PYR or L-alanine (ALA) [69].

In analogy to GC-EI-MS, the ^{13}C -distribution of several CID fragments may be combined. For LC-ESI-MS/MS fragments, this amounts to combining several TMID. The structural information gain with respect to the labeling state $\mathbf{x}$ can be nicely illustrated using the running example of the $[\text{M}+\text{H}]^+$ ion of ASP. For the fragment at m/z 74 of ASP in Figure 2.5, the carbon atoms of which stem from positions [1-2], the following measurement matrix can be derived:

$$\begin{pmatrix}
 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
 \end{pmatrix} \cdot \begin{pmatrix} x_{0000} \\ x_{0001} \\ x_{0010} \\ x_{0011} \\ x_{0100} \\ x_{0101} \\ x_{0110} \\ x_{0111} \\ x_{1000} \\ x_{1001} \\ x_{1010} \\ x_{1011} \\ x_{1100} \\ x_{1101} \\ x_{1110} \\ x_{1111} \end{pmatrix} = \begin{pmatrix} M0 > m0 \\ M1 > m0 \\ M1 > m1 \\ M2 > m0 \\ M2 > m1 \\ M2 > m2 \\ M3 > m1 \\ M3 > m2 \\ M4 > m2 \end{pmatrix} \quad (2.14)$$

Following the terminology in Tepper and Shlomi [171], we denote the vector on the right hand

side in equation (2.13) $\mathbf{TMID}_{\text{ASP},88}$ and in equation (2.14) $\mathbf{TMID}_{\text{ASP},74}$ to differentiate TMID of different fragments. The binary matrices in both equations are correspondingly denoted $\mathbf{C}_{\text{ASP}}^{88}$ and $\mathbf{C}_{\text{ASP}}^{74}$. Using LC-ESI-MS/MS technology both vectors $\mathbf{TMID}_{\text{ASP},88}$ and $\mathbf{TMID}_{\text{ASP},74}$ are given as numbers. To compute the isotopomer vector $\mathbf{x}_{\text{ASP}}$ from $\mathbf{TMID}_{\text{ASP},88}$ or $\mathbf{TMID}_{\text{ASP},74}$, the matrices $\mathbf{C}_{\text{ASP}}^{88}$ and $\mathbf{C}_{\text{ASP}}^{74}$ need to have full rank of 16. It turns out that $\mathbf{C}_{\text{ASP}}^{88}$ has a rank of 8 and $\mathbf{C}_{\text{ASP}}^{74}$ has a rank of 9. Thus, neither equation (2.13) nor equation (2.14) can be inverted for the isotopomer vector $\mathbf{x}_{\text{ASP}}$. However, the TMID of both fragments can be combined in the following way:

$$\begin{pmatrix} \mathbf{C}_{\text{ASP}}^{88} \\ \mathbf{C}_{\text{ASP}}^{74} \end{pmatrix} \cdot \begin{pmatrix} x_{0000} \\ x_{0001} \\ x_{0010} \\ x_{0011} \\ x_{0100} \\ x_{0101} \\ x_{0110} \\ x_{0111} \\ x_{1000} \\ x_{1001} \\ x_{1010} \\ x_{1011} \\ x_{1100} \\ x_{1101} \\ x_{1110} \\ x_{1111} \end{pmatrix} = \mathbf{C}_{\text{ASP}}^{88,74} \cdot \mathbf{x}_{\text{ASP}} = \begin{pmatrix} \mathbf{TMID}_{\text{ASP},88} \\ \mathbf{TMID}_{\text{ASP},74} \end{pmatrix} \quad (2.15)$$

The matrix arising from concatenating $\mathbf{C}_{\text{ASP}}^{88}$ and $\mathbf{C}_{\text{ASP}}^{74}$ in equation (2.15) has rank of 12. Thus, both TMID in conjunction are more informative with respect to $\mathbf{x}_{\text{ASP}}$ than each single TMID, but still do not allow to infer $\mathbf{x}_{\text{ASP}}$ outright. The process of concatenating individual measurement matrices of type $\mathbf{C}_{\text{ASP}}^{88}$ can be continued to include the TMID of further fragments with unique carbon origin. This was accomplished by Choi and Antoniewicz [29] using GC-EI-MS/MS. The authors used a combination of EI fragmentation and subsequent CID of precursor ions to acquire the TMID of several fragments of ASP concurrently. By combining the obtained TMID, the authors succeeded in elucidating the entire isotopomer distribution of ASP, the inclusion of which into the flux estimation improved statistical flux identifiability [28].

5.3.6 Data Acquisition

If the increased isotopomer resolution is to be harvested, the TMID data of as many fragments as analytically possible need to be acquired and interpreted. Currently, this faces two challenges, the first of which concerns product ion quantification. Quadrupole technology for product ion detection

imposes limits on the maximal number of fragment ion mass traces that can be acquired with a given accuracy (determined by cycle time) and sensitivity (influenced by dwell time) due to the interdependency of dwell time, cycle time and number of mass transitions. This issue was addressed by Rühl et al. [156] for LC-ESI-QqQ by setting the Q3 transmission windows such that all isotopomers contributing to a given mass isotopomer are transmitted or by setting the Q1 transmission windows to let pass all mass isotopomers contributing a certain tandem mass isotopomer. The inherently lower acquisition rate of a scanning quadrupole device may also be circumvented by restricting the number of acquired product ion spectra to the mass isotopomers surpassing an intensity threshold in an information-dependent acquisition as employed by McCloskey et al. [120]. Alternatively, the entire product ion spectrum of a given mass isotopomer may be acquired at once using a non-scanning mass analyzer like TOF MS as demonstrated by Mairinger et al. [115] for GC-Chemical Ionization (CI)-QqTOF.

5.3.7 Interpretation of Tandem Mass Isotopomer Distributions

The second challenge concerns the correct interpretation of the acquired TMIDs. The information gain outlined in equation (2.15) can only be harnessed based on the knowledge of the positional origin of fragment carbon atoms from the precursor ion. This knowledge may be inferred from annotating the corresponding mass peaks with a structure. Whereas LC-MS/MS spectra of peptides are easily predicted and annotated with structures due to straightforward fragmentation rules, the prediction of tandem MS spectra of small molecules poses a veritable challenge [198].

For structural elucidation the above mentioned studies resorted to fragment spectrum prediction using heuristic fragmentation rules as implemented in MassFrontier or combinatorial fragmentation as implemented in the proprietary software PeakView or openly available software FiD [71]. The latter approach exhaustively enumerates all fragment structures as sub-graphs of the precursor ion structure matching the observed (accurate) mass [80]. However, these approaches will fail to predict cyclic structures or product ions derived thereof because computational approaches to predict fragment structures do not include possible rearrangement reactions except for hydrogen to keep the dimensionality of the fragment space low enough for the problem to be computationally feasible [71]. The latest development for the computational prediction of MS/MS spectra is termed Competitive Fragmentation Modeling and is based on a machine-learning model trained on mass spectra of known structures [5]. The sophistication notwithstanding, this approach will not be able to capture cyclic fragment structures because the space of admissible structures is as restrictive as in the established computational approaches [6]. Cyclic product ion structures, though, occur frequently for amino acids in positive ionization mode [155].

Furthermore, the studies of Rühl et al. [156] and McCloskey et al. [120] left the product ion structures for many fragment ions of central carbon metabolism open, not least because several alternative fragment structures may explain a given elemental composition. This is especially true for fragments of di- and tricarboxylic acids corresponding to a neutral loss of carbon dioxide. If different carboxyl

groups are lost simultaneously, the resulting mass peak is composed of a mixed carbon ancestry. This has been shown to affect the fragment of CIT used for ^{13}C -measurement in the study of Alves et al. [7]: the used fragment is symmetrically derived from two positions within the precursor molecule. Although this study is the first to use LC-ESI-MS/MS data with experimentally validated carbon origins, it falls short of a comprehensive characterization of carbon atom fate over the central carbon metabolism, reporting the carbon ancestry for just one fragment of GLU, CIT and ASP each.

5.3.8 Flux estimation

The fifth pillar of ^{13}C -MFA concerns the numerical procedure to infer fluxes. This procedure relies on the assumption of constant pools sizes, i.e. metabolic steady state. At metabolic steady state, metabolite balancing yields a linear homogenous equation system comprising the stoichiometric matrix $\mathbf{N}$ and the flux vector $\mathbf{v}$. The flux vector contains forward as well as backward fluxes. Some fluxes of $\mathbf{v}$ may be assumed to be known from extracellular rate measurements. The mapping of the respective fluxes onto their measurement value is always linear and can therefore be described by the flux measurement matrix $\mathbf{M}_v$. The actual flux measurements are contained in vector $\mathbf{w}$. Normally, they come with an error estimate or are biased in the worst case. However, for the outline of the flux estimation problem $\mathbf{w}$ is assumed to be noise-free. The stoichiometric matrix and the flux measurement matrix can be put into one equation system, yielding:

$$\begin{pmatrix} \mathbf{N} \\ \mathbf{M}_v \end{pmatrix} \cdot \mathbf{v} = \begin{pmatrix} \mathbf{0} \\ \mathbf{w} \end{pmatrix} \quad (2.16)$$

This system does not admit an unique solution for $\mathbf{v}$ if the rank of the concatenated matrix is smaller than the dimensionality of $\mathbf{v}$. By introducing the free fluxes $\mathbf{v}_f$ [193], the flux space is parametrized according to:

$$\mathbf{v} = \mathbf{V} \cdot \mathbf{v}_f + \mathbf{v}_0 \quad (2.17)$$

Here $\mathbf{V}$ is the kernel matrix and $\mathbf{v}_0$ represents a particular solution of equation system (2.16). The free fluxes represent the missing flux information and may be inferred from fitting an isotopomer model of the metabolic network in question to ^{13}C -labeling measurements. An isotopomer model under metabolic steady state is generally given as a system of nonlinear ordinary differential equations of the following form [135]:

$$\begin{aligned} \text{diag}(\mathbf{c}_M) \frac{d\mathbf{x}(t)}{dt} &= \mathbf{f}(\mathbf{x}(t), \mathbf{x}_{\text{inp}}, \mathbf{v}), \\ \mathbf{g}(\mathbf{x}(t)) &= \mathbf{y}(t), \\ \mathbf{h}(\mathbf{c}_M) &= \mathbf{z} \end{aligned} \quad (2.18)$$

Here $\mathbf{f}(\mathbf{x}(t), \mathbf{x}_{\text{inp}}, \mathbf{v})$ is linear in the fluxes $\mathbf{v}$ and nonlinear (polynomial) in the labeling state vector $\mathbf{x}(t)$. $\mathbf{x}(t)$ represents the complete isotopic labeling state of the metabolic network which, depending on the modeling framework, contains isotopomers [160], EMUs [10] or cumomers [195]. The subsequent analysis employs the cumomer balancing approach. $\mathbf{c}_M$ denotes a matrix with concentrations of balanced intracellular metabolite pools on its diagonal with one dimension for each balanced cumomer. For a detailed description of the cumomer modeling framework the reader is referred to [193, 194, 195]. The dynamic cumomer modeling framework is described in Nöh et al. [135], Nöh et al. [130]. $\mathbf{x}_{\text{inp}}$ denotes the constant vector of input labeling fractions (given by the substrate composition), whereas $\mathbf{y}(t)$ and $\mathbf{z}$ denote the vector of label and pool size measurements, respectively, which are again assumed to be noise-free for the sake of illustration.

In experiments which fulfill the condition of isotopic stationarity the time differential on the left hand side of equation (2.18) vanishes, making the solution for $\mathbf{x}$ independent of the pool size vector $\mathbf{c}_M$. Therefore, flux estimation under isotopic stationary conditions is concerned with an algebraic equation system of the following form:

$$\begin{aligned} \mathbf{f}(\mathbf{x}, \mathbf{x}_{\text{inp}}, \mathbf{V} \cdot \mathbf{v}_f + \mathbf{v}_0) &= \mathbf{0}, \\ \mathbf{g}(\mathbf{x}) &= \mathbf{y} \end{aligned} \quad (2.19)$$

Here $\mathbf{v}$ has been substituted by its parameterization from equation (2.17). The flux estimation problem in general is concerned with finding a solution for $\mathbf{v}_f$ from equation (2.19), given the vector of measured labeling enrichments $\mathbf{y}$. Since this vector is experimental and therefore noisy and most likely systematically biased, equation system (2.19) cannot be simply inverted for $\mathbf{v}_f$. Instead, a non-linear regression approach needs to be taken, reconciling all available data. Using the employed substrate isotopomer mixture, it is possible to compute the ^{13}C -label distribution for one assumed setting of free fluxes through a cascade of linear equations [195]. Through iterating the assumed free fluxes, the difference between measured label and computed label is minimized. The objective function for this optimization problem can be given as as:

$$\underset{\mathbf{v}_f}{\operatorname{argmin}} \sum_i^n \left(\frac{(y_i(\mathbf{x}(\mathbf{v}_f, \mathbf{x}_{\text{inp}})) - \hat{y}_i)}{\hat{\sigma}_{y_i}} \right)^2 + \sum_j^m \left(\frac{w_j(\mathbf{v}_f) - \hat{w}_j}{\hat{\sigma}_{w_j}} \right)^2 \quad (2.20)$$

In equation (2.20) hats are used to differentiate real from computed measurements y_i for one given isotopomer vector. Each difference is weighted by an estimate of the standard deviation of the measurement. To solve equation (2.20) a suite of software package have been developed, most prominently among them Metran [202], OpenFLUX [151] and 13CFLUX2 [188]. The latter distinguishes itself in using a unified specification of carbon atom transition models in form of xml-files, in analogy to the systems biology markup language. It furthermore incorporates many built-in functionalities, such as support for MS/MS data. The computation of TMID distributions needs to be performed manually post-simulation in OpenFLUX [151]. Equipped with functions to uniformly sample free

fluxes from the polyhedron describing the constrained nullspace, the 13CFLUX2 package supports thorough multistart optimizations.

5.3.9 Plausibility Control and Data Visualization

Once a flux distribution or set of flux distributions satisfying the criterion of minimal difference between simulated and experimental ^{13}C -distributions are obtained, the result must be tested for biological plausibility. It has to be born in mind that the obtained solution is simply the one reconciling all measurements. As such, the underlying flux distribution may not concordant with the obtained solution. If the researcher trusts the solution, nonetheless, the result can be visualized. Here, public software solutions are far less developed than those for simulation. OMIX is the sole dedicated and professionally developed software suite for this purpose [\[44\]](#).

Materials and Methods

6 Materials

6.1 Strain maintenance

The strains investigated in this thesis are given in Table 4.3 and were kindly provided by Bastian Blomblach, Institute for Biochemical Engineering, University of Stuttgart. Each strain arrived at the Institute IBG-1 on agar plates. Upon arrival one colony was picked from each plate with a sterile wooden toothpick. One part was transferred to a PCR tube for colony PCR analysis to confirm strain identity. Subsequently, a shaking flask with CGXII preculture medium containing 1% glucose (w/v) was inoculated with the corresponding strain by dropping the wooden stick into it. Each culture was harvested during the late exponential phase by centrifuging it at 4500 rpm for 10 min, resuspending it once in sterile saline and finally resuspending it in 20 % (v/v) glycerol solution in sterile 0.9 % (w/v) NaCl in distilled water. From this solution cryo stocks were prepared and immediately stored at -80°C.

The Δ (PCx PEPCx) strain was picked into BHI medium. After growing to stationary phase, the culture was used to inoculate defined CGXII medium with 0.8% (w/v) sodium acetate. From this culture, CGXII medium with 2% (w/v) sodium acetate as carbon source was inoculated. From this culture, the final cryo culture was made as described above.

6.2 Chemicals

If not otherwise stated, all non-labeled chemicals were acquired from Sigma Aldrich (Schnelldorf, Germany). Acetyl-CoA trilithium salt was acquired from Roche Diagnostics (Mannheim, Germany). Succinic semialdehyde was purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA). (R)-3-Aminodihydrofuran-2,5-dione hydrochloride was purchased from Interchim (Montlucon, France). Trans-aconitic anhydride was synthesized by FCH Group (Chernigiv, Ukraine) and acquired through AKOS GmbH (Steinen, Germany). Citric anhydride was synthesized according to Repta and Higushi [153]. Acetic anhydride and acetic acid for the synthesis of citric anhydride were acquired from Roth (Karlsruhe, Germany).

6.3 Isotopic standards

All stable ^{13}C -isotopes had an atomic purity of 99 %. $\text{H}_2[^{18}\text{O}]$ was acquired from Sigma Aldrich and had an ^{18}O -enrichment of 10 %. $\text{NaH}[^{13}\text{C}]\text{O}_3$ was acquired from MSD Isotopes (Montreal, Canada). Sodium $[2\text{-}^{13}\text{C}]\text{pyruvate}$ and sodium $[3\text{-}^{13}\text{C}]\text{pyruvate}$ were purchased at Isotech (Miamisburg, OH, USA). Sodium $[1\text{-}^{13}\text{C}]\text{pyruvate}$, $[1\text{-}^{13}\text{C}]\text{L-glutamic acid}$ and sodium $[1,2,3,4\text{-}^{13}\text{C}]\alpha\text{-ketoglutarate}$ were acquired from Cambridge Isotope Laboratories (Andover, MA, USA). $[4\text{-}^{13}\text{C}]\text{L-aspartic acid}$, $[1\text{-}^{13}\text{C}]\text{L-aspartic acid}$, $[5\text{-}^{13}\text{C}]\text{L-methionine}$ and $[6\text{-}^{13}\text{C}]\text{L-lysine dihydrochloride}$ were purchased from Sigma Aldrich (Schnelldorf, Germany). $[2\text{-}^{13}\text{C}]\text{Glucose}$, $[1,2\text{-}^{13}\text{C}]\text{Glucose}$, $[6\text{-}^{13}\text{C}]\text{Glucose}$ were acquired from

6. MATERIALS

Omicron Biochemicals (South Bend, IN, USA) and [1-¹³C]Glucose from Campro Scientific (Berlin, Germany).

[¹⁵N]Asp was acquired from Sigma Aldrich (Schnelldorf, Germany). ¹⁵N-labeled and naturally labeled Maltose binding protein from *Escherichia coli* was acquired from Giotto Biotech S.r.l. (Florence, Italy). The protein lacks the signal peptide (residues 1-26) and four trailing amino acids and therefore comprises residues 27-392 of the sequence deposited under accession number P0AEX9 on Swiss Prot.

6.3.1 Enzymes

The assay enzymes pyruvate carboxylase (bovine liver), L-glutamic dehydrogenase (bovine liver), citrate synthase (porcine heart), aconitase (porcine heart), phosphoglucose isomerase (*Saccharomyces cerevisiae*), fructose-6-phosphate kinase (*Bacillus stearothermophilus*) and glucose-6-phosphate dehydrogenase (*Leuconostoc mesenteroides*) were acquired from Sigma Aldrich (Schnelldorf, Germany); malate dehydrogenase (porcine heart) and hexokinase (*Saccharomyces cerevisiae*) were purchased at Roche Diagnostics GmbH (Mannheim, Germany).

6.4 Consumable solutions and chemicals

Table 3.1: Consumable chemicals, solutions and non-standard materials.

Solution or chemical	Manufacturer	Product Number
APCI Negative Calibration Solution for the AB SCIEX TripleTOF Systems	AB SCIEX Darmstadt, Germany	4460134
APCI Positive Calibration Solution for the AB SCIEX TripleTOF Systems	AB SCIEX Darmstadt, Germany	4460131
CASY ton	OMNI Life Science GmbH Bremen, Germany	5651808
3 μ m latex beads size calibration standard	Beckman Coulter Krefeld, Germany	6602793
0.1 % formic acid in water ULC/MS	Biosolve BV Valkenswaard, The Netherlands	0023244102BS
Acetonitrile LC-MS	Biosolve BV Valkenswaard, The Netherlands	001207802BS
Methanol absolute LC-MS	Biosolve BV Valkenswaard, The Netherlands	0013687802BS
Tributylamine puriss. plus	Sigma Aldrich Schnelldorf, Germany	90781
Acetic acid analytical standard	Sigma Aldrich Schnelldorf, Germany	71251
Ammonium acetate p.a.	Sigma Aldrich Schnelldorf, Germany	32301
Ammonia 25 % p.a.	ROTH Karlsruhe, Germany	67742
Antifoam 204	Sigma Aldrich Schnelldorf, Germany	A6426
Flowerplate BOH	M2P-labs Baesweiler, Germany	MTP-48-BOH

6. MATERIALS

6.5 Media

Table 3.2: Defined CGXII medium for bioreactor cultivations.

(a) CGXII medium according to Unthan et al. [177]		(b) stock solutions for CGXII medium	
Component	Amount per liter	Component	Concentration
NH ₄ SO ₄	20 g	trace element stock solution	
KH ₂ PO ₄	1 g	FeSO ₄ * 7 H ₂ O	10 g l ⁻¹
K ₂ HPO ₄	1 g	MnSO ₄ * H ₂ O	10 g l ⁻¹
CaCl ₂ stock solution	1 ml	ZnSO ₄ * 7 H ₂ O	1 g l ⁻¹
MgSO ₄ stock solution	1 ml	CuSO ₄ * 5 H ₂ O	0.313 g l ⁻¹
ad. 0.75 l distilled water		NiCl ₂ * 6 H ₂ O	0.02 g l ⁻¹
Adjust pH to 7 with 4 M NaOH		adjust to pH=1 with 4M HCL	
autoclave at 121°C, 2 bar for 20 minutes		calcium chloride stock solution	
trace element stock solution	1 ml	CaCl ₂ * 2 H ₂ O	13.25 g l ⁻¹
biotin stock solution	1 ml	magnesium sulfate stock solution	
protocatechuate stock solution	2 ml	MgSO ₄ * 7 H ₂ O	250 g l ⁻¹
glucose stock solution	25 ml	protocatechuate stock solution	
antifoam 204 undiluted	0.2 ml	protocatechuic acid	15 g l ⁻¹
ad. 1 L distilled water		Adjust to pH=9 with 4M NaOH	
		biotin stock solution	
		biotin	0.2 g l ⁻¹
		Heat up to 45°C to dissolve	
		glucose stock solution	
		glucose * 1 H ₂ O	440 g l ⁻¹

Table 3.3: CGXII media for non-pH-regulated cultivations systems.

(a) CGXII medium for precultures		(b) CGXII medium for BioLector cultivations	
Component	Amount per liter	Component	Amount per liter
NH ₄ SO ₄	20 g	NH ₄ SO ₄	20 g l
KH ₂ PO ₄	1 g	KH ₂ PO ₄	1 g l
K ₂ HPO ₄	1 g	K ₂ HPO ₄	1 g l
MOPS	42 g	Urea	5 g l
CaCl ₂ Stock	1 ml	MOPS	42 g l
MgSO ₄ Stock	1 ml	CaCl ₂ Stock	1 ml
ad. 0.75 l distilled water		MgSO ₄ Stock	1 ml
Adjust pH to 7 with 4 M NaOH		ad. 0.75 l distilled water	
autoclave at 121°C, 2 bar for 20 minutes		adjust pH to 7 with 4 M NaOH	
trace element stock solution	1 ml	autoclave at 121°C, 2 bar for 20 minutes	
biotin stock solution	1 ml	trace element stock solution	1 ml
protocatechuate stock solution	2 ml	biotin stock solution	1 ml
glucose stock solution	25 ml	protocatechuate stock solution	2 ml
		glucose stock solution	25 ml

6.6 Devices and Chromatographic Material

Table 3.4: List of devices used in this work.

Device	Manufacturer	Application
parallel cultivation system	DASGIP Jülich, Germany	parallel magnet-stirred bioreactor cultivation of <i>C. glutamicum</i>
INFORS HT incubation shaker	Infors Bottmingen, Switzerland	preculture of <i>C. glutamicum</i>
API4000 QqQ MS	AB SCIEX Darmstadt, Germany	targeted quantification of primary intermediates
TripleTOF 6600 QqTOF MS	AB SCIEX Darmstadt, Germany	SWATH acquisition and full spectral ^{13}C -profiling
4000 QTRAP QLIT MS	AB SCIEX Darmstadt, Germany	targeted proteomics of <i>C. glutamicum</i>
Multisizer3 Coulter Counter	Beckman Coulter Krefeld, Germany	equipped with 30 μM aperture tube for cell size distribution measurement
Agilent 1200 HPLC system	Agilent Technologies Waldbronn, Germany	HPLC separation prior to QqQ MS detection of primary metabolites
Agilent 1100 HPLC system	Agilent Technologies Waldbronn, Germany	equipped with refractive index detector for detection of glucose in supernatant samples
Agilent 1260 Infinity HPLC system	Agilent Technologies Waldbronn, Germany	HPLC separation prior to QqTOF MS detection of primary metabolites or peptides
UV-1800 UV-Spectrophotometer	Shimadzu Duisburg, Germany	rapid optical biomass quantification
Reak 2 Overhead shaker	Heidolph Schwabach, Germany	cell lysis and extraction for intracellular metabolome analysis
Allegra 64R Centrifuge	Beckman Coulter Krefeld, Germany	cell separation at -20°C for intracellular metabolome analysis
GS-15R Centrifuge	Beckman Coulter Krefeld, Germany	cell separation at 20°C for washing of cells
CX4000 Gas Analyzer	GASMET Technologies Oy Helsinki Finland	discrimination of CO_2 and $^{13}\text{C}[\text{O}_2]$ during ^{13}C -labeling experiments
BioLector micotiter cultivation system	M2P-labs Baesweiler, Germany	micro-scale cultivation of <i>C. glutamicum</i>

Table 3.5: List of HPLC columns used in this work.

Commercial Name	Packing material	Particle size [μm]	Length * internal diameter [mm * mm]	Product, catalogue or part number
SeQuant ZIC-pHILIC Main Column (Merck KGaG, Darmstadt, Germany)	PEEK polymer	5	150 * 2.1	1.50460.0001
SeQuant ZIC-pHILIC Guard Column (Merck KGaG, Darmstadt, Germany)	PEEK polymer	5	20 * 2.1	1.50437.0001
C18 synergy hydro Reversed Phase (Phenomenex, Torrance, CA, USA)	silica	4	150 * 2	00F-4375-B0
Ascentis Express Peptide ES-C18 (Sigma Aldrich, Schnelldorf, Germany)	silica	2.7	150 * 2.1	53307-U
Acclaim PepMap Trap Cartridge (Thermo Scientific, Bremen, Germany)	silica	5	5 * 0.3	163589
Luna SCX (Phenomenex Torrance, CA, USA)	silica	5	150 * 2	00F-4398-B0
CS Organic Acid Resin Main Column Refill (CS Chromatographie, Langerwehe, Germany)	PS-DVB polymer	10	300 * 8	52998030
CS Organic Acid Resin Precolumn (CS Chromatographie, Langerwehe, Germany)	PS-DVB polymer	10	40 * 8	548980

6.7 Software

Table 3.6: List of software used for the present work.

Software name	Distributor	Application
MATLAB R2013b and R2016b	MathWorks	Data analysis, optimization, and visualization
Mathematica 8.0	Wolfram	Integer computations, and determination of feasible elemental compositions
13CFLUX2	IBG-1, Forschungszentrum Jülich	Flux estimation from ^{13}C -label data
HQQ	IBG-1, Forschungszentrum Jülich	Integration of XIC from QqQ MS-based acquisition methods
PeakView 2.1	ABSCIEX	Evaluation of QqTOF MS mass spectra
MultiQuant 3.0.1	ABSCIEX	Integration of XIC from QqQ MS-based and QqTOF MS-based acquisition methods
ChemDraw Professional 15.0.0.106	PerkinElmer	Chemical structure typesetting
CorelDRAW X6	Corel Corporation	Visualization
Analyst 1.6.2	ABSCIEX	Control of QqQ MS and analysis of QqQ MS data
Analyst TF 1.7	ABSCIEX	Control of QqTOF MS
ProteinPilot 4.5	ABSCIEX	Identification of peptides against FASTA database
OMIX 1.9.33	OMIX Visualization GmbH	Visualization of metabolic networks

7 Methods

7.1 Bioreactor cultivations

Bioreactor cultivations were carried out in a parallel fermentation system (DASGIP, Jülich, Germany). The composition of the medium for bioreactor cultivations is given in Table 3.2a. The assembled

bioreactors were autoclaved with 3 ml of distilled water filled into the system. After sterilization, the vessels were sterilely filled with 200 ml of medium in a safety cabinet.

A preculture in shaking flasks with defined CGXII medium as described in Table 3.3a was inoculated from cryo-culture and incubated overnight at 30°C and 250 rpm in an incubation shaker. The preculture medium had the same carbon composition as the medium used in the bioreactor cultivation. At the time of harvest, the optical density of the preculture was measured. Precultures with optical densities in the range of $OD_{600} = 3-5$ were used to inoculate bioreactors. To reach an inoculation density of $OD_{600} = 0.5$ in 200 ml of bioreactor working volume, a total of $100 / OD_{600}$ ml of preculture broth was sterilely transferred into a falcon tube. The preculture was centrifuged for 10 minutes, 4500 rpm in a GS-15R centrifuge (Beckman Coulter, Krefeld, Germany), washed once with sterile 0.9 % NaCl solution and again centrifuged. The pellet was finally resuspended in 2 ml of sterile 0.9 % NaCl solution. This volume was aspirated with a syringe and injected into the bioreactor through a septum sterilized with 70 % ethanol.

During bioreactor cultivations dissolved oxygen (DO) levels were maintained above 30 % by adjusting stirrer speed and oxygen content of the inlet air. The gassing rate was set to 1 vvm [volume of air per volume liquid per minute in $l (l \text{ min})^{-1}$] in 200 ml working volume except otherwise stated. The pH was maintained at pH 7 by feeding either 4M NaOH or 4M HCl. The cultivations temperature was 30°C.

7.2 Metabolome analysis

Cultivation experiments for the metabolome analysis were carried out identically to the bioreactor cultivations described in subsection 7.1, except that each reactor was equipped with a sampling tube. Through this tube culture samples were drawn into a syringe in technical duplicate at two time points. These time points correspond to target biovolume (BV) concentrations of 5 and 10 $\mu l \text{ ml}^{-1}$, respectively, yielding a total of four technical replicates per strain. To obtain a target total BV of 50 μl in the sample, the sample volume to be quenched, V_{Sample} , was 10 and 5 ml for the two sampling time points, respectively.

From the drawn volume of culture broth, the sample volume V_{Sample} was pipetted with a Research plus pipette (Eppendorf, Hamburg, Germany) into the quenching solution consisting of 60 % methanol in MilliQ water at approximately -50°C. The volume of quenching solution V_{Quench} was three times the sample volume to be quenched and aliquoted at room temperature and frozen at -80°C in 50 ml falcon tubes upfront. The resulting solution of quenching solution and culture broth was shortly agitated and immediately centrifuged at -20°C for 10 minutes at 4500 rpm. After centrifugation, some part of the 'quenching supernatant' was filtered through a 0.2 μm cellulose acetate filter and immediately stored at -20°C. After the remaining quenching supernatant was decanted, the falcon tube was placed in a metal rack pre-cooled at -20°C and turned upside down onto a paper towel. In this position the tube was held for 30 s to let run out the remaining quenching supernatant. Afterward the tubes were immediately frozen in metal racks at -80°C.

7. METHODS

Parallel to the quenching, a part of the culture sample was filtered through a 0.2 μm cellulose acetate filter. 250 μl of this 'culture supernatant' was pipetted to 750 μl of 60 % methanol in MilliQ at -20°C . This volume was pipetted upfront at room temperature. A third part of the drawn culture volume was used to measure the BV concentration according to section 7.4. This BV concentration measurement together with the sampled culture volume was used to calculate the BV contained in the pellet of each sample.

During the following extraction, the falcon tubes containing the biomass pellet were placed into a liquid ethanol bath maintained at a temperature of at most -27°C . The extraction was carried out by first adding to each sample a volume of chloroform at -20°C equivalent to 50 times the BV of the sample. Subsequently, a volume of extraction solvent V_{Ext} (50 % TE buffer at pH 8 and 50 % methanol at -20°C) equivalent to 50 times the BV was added. Each sample was shortly vortexed before being incubated at -20°C in an overhead shaker for at least two hours. Afterward each sample was centrifuged at -20° for 10 minutes at 4500 rpm. The upper methanolic phase from the resulting three phases was aspirated with caution into a pipette tip and filtered through a 0.2 μm cellulose acetate filter. This sample constitutes the intracellular extract sample.

For LC-MS/MS analysis 50 μl of each extract, quenching supernatant and culture supernatant sample was spiked with 50 μl of internal standard. The internal standard for the analysis of sugar phosphates and organic acids was an extract from *E. coli* grown on fully ^{13}C -labeled glucose. For amino acids the internal standard was a quenching supernatant from a cultivation of *C. glutamicum* on fully ^{13}C -labeled glucose.

Concentration of organic acids and sugar phosphates were calibrated by an external standard mixture ranging from 0.1 μM to 100 μM , spiked with the same identical standard as the samples. Depending on the linearity of the $^{12}\text{C}/^{13}\text{C}$ ratio, higher concentrations were excluded from the linear regression to obtain a R^2 of at least 0.98. For amino acids, a standard series ranging from 0.025 μM to 25 μM was spiked with the internal standard used for the samples. Dilutions of the samples were necessary for some amino acids and were performed with 50 % methanol in MQ water at -20°C . The samples were also maintained at -20°C . Integration of mass traces for metabolites and internal standards was performed with the in-house software HQQ.

To calculate the intracellular concentrations, leakage effects need to be corrected for, depending on the metabolite. To judge if leakage occurred for one particular metabolite, the difference between its measured concentration in the quenching supernatant and the culture supernatant sample was computed. This difference is expected to be positive for any metabolite which is subject to leakage effects, zero if no leakage occurred and negative if concentration measurements were biased. A two-sided t-test was performed on the concentration differences at a significance level of 0.05 with 31 degrees of freedom since 32 (eight strains * four replicates) replicates were available. The decision on whether leakage occurred or not was pooled over all samples from all strains assuming that the causes of this phenomenon are not different among the investigated strains.

When leakage occurred, the amount of leaked metabolite in one sample was calculated according

to:

$$n_{\text{Leakage}} = n_{\text{QS}} - n_{\text{CS}} = c_{\text{QS}}(V_{\text{Sample}} + V_{\text{Quench}}) - c_{\text{CS}} V_{\text{Sample}} \quad (3.1)$$

where the subscript acronyms QS and CS denote 'quenching supernatant' and 'culture supernatant', respectively. The value of c_{CS} in equation (3.1) already accounts for the 1:4 dilution of the culture supernatant. When no leakage occurred, n_{Leakage} was zero. The final concentration of a metabolite in one sample can be then calculated by normalizing the amount of metabolite extracted from the sample by the total BV according to:

$$c_{\text{intra}} = \frac{n_{\text{Leakage}} + c_{\text{Ext}} V_{\text{Ext}}}{BV_{\text{sample}}} \quad (3.2)$$

where c_{Ext} is the measured concentration in the extract sample, and V_{Ext} is the used volume of the extraction solvent (50 % TE buffer (pH 8) and 50 % methanol) in the sample. Equation (3.2) was used to calculate the concentration of each measured metabolite in each of the four replicates per strain.

7.3 BioLector cultivations

BioLector cultivations were performed in the BioLector parallel cultivation system (m2p-labs, Baesweiler, Germany). Flowerplates with optodes for optical pH and DO measurements were employed. The cryo-culture for inoculation was washed once with sterile saline and resuspended in glucose-free CGXII medium. From this incolum 50 μl were transferred into each well containing 950 μl 1 % glucose medium. The plate was sterilely sealed and incubated at 30°C an 1300 rpm.

7.4 Analytics for extracellular phenotyping

Optical density

Optical density (OD) was measured at 600 nm in a UV-1800 UV-spectrophotomoer (Shimadzu, Duisburg, Germany). Dilutions were performed in 0.9 % (w/v) NaCl. Dilutions were performed to reach an optical density between 0.1 and 0.3.

Biovolume

Cultivation samples were diluted 1:200 or 1:2000 depending on the biomass concentration in 10 ml CASY ton Puffer (OMNI Life Science GmbH, Bremen, Germany). The size distribution of the sample was determined by the MultiSizer3 Coulter Counter (Beckman Coulter, Krefeld, Germany). The BV concentration was computed by calculating the first moment of the distribution and assuming a spherical shape of the measured cells. The resulting quantity has the unit $\mu\text{l ml}^{-1}$.

7. METHODS

Cell dry weight

Cell dry weight (CDW) was determined by centrifuging 2 ml of a bioreactor sample in a pre-dried and pre-weighted Eppendorf tube at 13,000 rpm for 7 min. The cells were washed in 1 ml 0.9 % (w/v) NaCl by resuspension and renewed centrifugation. To resuspend the pellet, special care was taken to avoid abrasion of the Eppendorf tube by protective sheets of paper. After decanting the supernatant the pellets were dried for at least 2 days at 80 °C. A correlation between BV concentration in $\mu\text{l ml}^{-1}$ and CDW concentration in g l^{-1} was derived for the $\Delta(\text{PCx ODx})$, $\Delta(\text{PEPCx ME})$ mutants and the WT and then applied to calculate the CDW for all cultivations assuming a standard deviation of 5%. The obtained correlation is shown in Appendix 2.

Acetate concentration (HPLC)

HPLC analytics were performed on an Agilent 1200 system (Agilent Technologies, Waldbronn, Germany) using CS Organic Acid Resin (CS Chromatographie Service GmbH, Langerwehe). Separation was performed isocratically with 0.1 M H_2SO_4 as mobile phase at a flow rate of 0.5 ml/min at 40°. Acetate and other organic acids were determined with a diode array detector UV detector measuring between 190-400 nm.

Glucose concentration (HPLC)

HPLC analytics for glucose quantification were performed on an Agilent 1110 system (Agilent Technologies, Waldbronn, Germany) using CS Organic Acid Resin (CS Chromatographie Service GmbH, Langerwehe). Separation was performed isocratically with 0.1 M H_2SO_4 as mobile phase at a flow rate of 0.5 ml min⁻¹ at 40°C. Glucose was detected by means of a refractive index detector maintained at 35°C.

Detection of CO₂ in exhaust gas

Exhaust gas analytics were performed with a gas analyzer module from DASGIP (DASGIP, Jülich, Germany) based on infrared absorption.

Detection of [¹³C]O₂ in exhaust gas

Detection of [¹³C]O₂ in the exhaust gas was performed by Fourier Transform Infrared Interference spectroscopy using a CX4000 gas analyzer (Gasmeter Technologies Oy, Helsinki, Finland). The gas analyzer was coupled to the outlet of the gas analyzer module of the DASGIP fermentation system.

Elemental Analysis of Biomass

The concentration of carbon, nitrogen and sulfur in biomass dry mass were determined at the central Analytical core facility of Forchungszentrum Jülich (ZEA-3). Biomass samples from the early stationary phase were processed following the same procedure as for CDW content determination. The dried biomass pellet was ground to a fine powder in a pre-dried mortar using a pre-dried pestle and sent in a sealed container to ZEA-3.

7.5 LC-ESI-QqQ analytics for central carbon metabolism intermediates

LC-ESI-QqQ method for amino acids

Samples for quantification and ^{13}C -isotopomer profiling of amino acids were separated using Luna SCX cation exchange column at 60°C on a JASCO HPLC system. The following buffers were employed: Buffer A: 5 % acetic acid, B: 15 mM ammonium acetate. The gradient elution of Table 3.7 was used for amino acid quantification as well as ^{13}C -profiling. The injection volume was $10\ \mu\text{l}$.

Table 3.7: Gradient elution for amino acid quantification and ^{13}C -profiling.

Time [min]	Flowrate [$\mu\text{l min}^{-1}$]	% B
0	400	15
10	400	15
16	400	100
28	400	100
30	400	15
35	400	15

LC-MS/MS method for organic acids

Samples for quantification and ^{13}C -isotopomer profiling of organic acids were separated using a synergy hydro C18 reversed phase column (see Table 3.5) on an Agilent 1200 chromatography system (Agilent Technologies, Waldbronn, Germany). The HPLC column outlet was coupled to a QqQ MS device (API 4000, AB Sciex, Darmstadt, Germany) equipped with a TurboSpray ion source in negative ionization mode. The elution was isocratic at 84 % buffer A and 16 % buffer B at a flowrate of $0.45\ \text{ml min}^{-1}$ at 20°C . The eluents were as follows: Buffer A: 10 mM tributylamine, 15 mM acetic acid, pH 4.95; Buffer B: methanol. MS parameters were as follows: CAD (collision gas pressure): 5, CUR (curtain gas flow): 30, GS1 (nebulizer gas flow): 70, GS2 (turbo

7. METHODS

heater gas flow): 70, IS (electrospray voltage): -4500 V, TEM (heater gas temperature): 650°C, entrance potential: -10 eV. Injection volume was 10 μ l.

LC-MS/MS method for sugar phosphates

Samples for quantification and ^{13}C -isotopomer profiling of sugar phosphates were separated on a synergy hydro C18 reversed phase column (see Table 3.5) at 40°C. Eluent were Buffer A: 10 mM tributylamine, 15 mM acetic acid, pH 4.95 and Buffer B: methanol. The HPLC system, the QqQ MS device and its MS parameter settings were the same as for the LC-MS/MS method for organic acids. For quantification the gradient of Table 3.8a was applied, whereas the gradient of Table 3.8b was used when ^{13}C -enrichment was measured. MS parameters were identical to the LC-MS/MS method for organic acids. Injection volume was 10 μ l.

Table 3.8: HPLC gradients for the elution of sugar and nucleoside phosphates for quantification and ^{13}C -enrichment profiling.

(a)	Gradient elution for sugar phosphate quantification.			(b)	Gradient elution for sugar phosphate ^{13}C -profiling.		
	Time [min]	Flowrate [$\mu\text{l min}^{-1}$]	% B		Time [min]	Flowrate [$\mu\text{l min}^{-1}$]	% B
	0	450	0		0	250	0
	2	450	0		6	250	0
	5	450	20		15	250	20
	8	450	20		24	250	20
	10	450	35		30	250	35
	14	450	100		42	250	100
	15	450	100		45	250	100
	15.5	450	0		46	250	0
	17	450	0		52	250	0

7.6 SWATH acquisition method

For SWATH acquisition of peptides samples were prepared according to the protocol of Voges and Noack [183] and separated on an Ascentis column preceded by a PepMap trap cartridge from Table 3.5. The HPLC methods employed 0.1 % formic acid and 0.1 % formic acid in acetonitrile (see Table 3.1) as buffers A and B, respectively, and consisted of the following gradient:

Table 3.9: HPLC method for SWATH Acquisition.

Total Time [min]	Flow Rate [$\mu\text{l min}^{-1}$]	% B
-12	200.00	3.0
0	200.00	3.0
70	200.00	40.0
78	200.00	40.0
79	200.00	60.0
89	200.00	60.0
90	200.00	3.0

Autosampler setting were 6°C and injection volume was 10 μl . The TripleTOF6600 was operated with CUR: 35, GS1: 30, GS2: 30, IS: 5500, TEM: 300, DP:120. During the elution under the above specified parameters the variable width Q1 windows given in Table 4 of Appendix 6 were monitored in a non-scheduled manner.

7.7 SWATH data processing

C. glutamicum WT was cultivated on CIT, MAL, benzoate, PYR, glucose and mixtures of xylose and glucose. Samples of each cultivation were processed as described in Voges and Noack [183]. An information-dependent acquisition was performed on an injection from these samples using the above described HPLC method. During the information-dependent acquisition all ions with m/z greater than 300, charge state 2-4 and above intensity of 150 were selected for fragmentation. Using ProteinPilot a library of confidently identified peptides was generated for each sample. All peptides were assembled in a meta-library of *C. glutamicum* covering 1727 proteins. This library incorporates the peptide confidence after identification, peptide (precursor) intensity in the MS1 scan from the information-dependent acquisition, fragment ion intensities and the observed peptide retention time.

For SWATH processing, the ProteinPilot library was imported into the MS/MS^{ALL} with SWATHTM Acquisition MicroApp within PeakView. For 12 fragments each of the ten most intense peptides (according to the MS1 precursor ion intensity from the IDA acquisition) the XIC were extracted from the SWATH spectra within an time interval of ± 8 minutes around the expected peptide retention time from the IDA-acquisition. Each ion trace was integrated and scored as implemented in the SWATHTM Acquisition MicroApp. Through extracting of a decoy sequence of each peptide and scoring the obtained fragment ion chromatograms analogously, a false discovery rate (FDR) can be calculated for each peptide. For the processing of the dataset presented in the Results and Discussion section, the false discovery rate threshold was set to 0.1 %. From all extracted 120 mass traces per protein only those with a false discovery rate less than the threshold value were exported and subjected to further

analysis.

7.8 Generation of product ion spectral libraries on a QqQ MS and QqTOF MS

The product ion spectra on the QqQ MS device were generated upon direct infusion of metabolite standards dissolved in 1:1 MilliQ water/methanol except for sugars of which a solution in ammonium acetate buffer at pH 9.2 was infused. Citric anhydride was injected as solution in 11% acetic anhydride in acetic acid (v/v). Each spectrum was the combination of 26 MCA scans recorded over a collision energy ramp from -130 eV to -5 eV for negative ionization mode and +5 eV to +130 eV for positive ionization mode at an optimized DP value. Collision gas was nitrogen (CAD parameter set to 5). Each spectrum is provided in Electronic Supplementary Material S7 of Kappelmann et al. [91].

Accurate mass product ion spectra were obtained by recording a high resolution product ion spectrum of the corresponding metabolites using an ESI-QqTOF (TripleTOF 6600, AB Sciex, Darmstadt, Germany) equipped with DuoSpray ion source. Direct infusion was performed as for the QqQ MS spectra. The data acquisition and peak annotation were performed using manufacturer software (Peakview 2.1, AB Sciex). QqTOF mass spectra were acquired for all considered precursor ions. Collision gas was nitrogen at 4 psi and the CE-ramp was recorded with an optimal DP-setting.

7.9 *In vitro* enzyme assays to generate selectively labeled ^{13}C -species and ^{18}O -species

To generate selectively labeled MAL and OAA species, a modified version of the pyruvate carboxylase assay of the manufacturer was employed [165]. Sodium [1- ^{13}C]pyruvate, sodium [3- ^{13}C]pyruvate or sodium pyruvate was incubated with either NaHCO_3 or $\text{NaH}[^{13}\text{C}]\text{O}_3$ to generate [1- ^{13}C]MAL, [3- ^{13}C]MAL and MAL or [1,4- ^{13}C]MAL, [3,4- ^{13}C]MAL and [4- ^{13}C]MAL, respectively. The assay volume was 1 ml and contained 98.5 mM triethanolamine-HCL (pH 8), 4.93 mM MgSO_4 , 6.45 mM sodium pyruvate, 0.026 mM acetyl-CoA, 0.99 mM Na_2ATP , 14.85 mM sodium bicarbonate, 0.228 mM Na_2NADH and 2.61 units/ml of malate dehydrogenase. For the generation of selectively labeled OAA species malate dehydrogenase and NADH were omitted. After the assay solution was allowed to thermally equilibrate at 30°C in a Thermomixer, the reaction was started by pipetting 3 μl of pyruvate carboxylase into the reaction tube, resulting in an enzyme concentration of 0.18 units/ml. After gentle agitation, the assay mixture was incubated for 30 minutes without agitation. After the successful completion was ascertained by measuring the absorbance at 320 nm, the assay solution was 0.2 μm -filtrated and immediately stored at -20°C. For LC-ESI-MS/MS analysis the assay samples were diluted with methanol to a concentration of 45 % (v/v) methanol and stored at -80°C till analysis.

To generate selectively labeled αKG a modified version of glutamate dehydrogenase assay of the manufacturer was employed [167]. The assay volume was 1 ml and contained 7.62 mM of either GLU or [1- ^{13}C]GLU and 0.015 units glutamate dehydrogenase per ml. The reaction was carried out at room temperature. After 60 minutes of incubation, successful completion was checked photometrically at 320 nm. All subsequent steps were carried out as described above.

Selectively labeled species of glucose 6-phosphate (G6P), fructose 6-phosphate (F6P), fructose 1,6-bisphosphate (FBP) and 6-phosphogluconate (6PG) were produced by incubating 200 μ M of either [1- 13 C]glucose, [2- 13 C]glucose, [1,2- 13 C]glucose or [6- 13 C]glucose with one or a combination of the enzymes hexokinase (at 5 units/ml), phosphoglucose isomerase (at 2 units/ml), fructose 6-phosphate kinase (at 3 units/ml) and glucose 6-phosphate dehydrogenase (at 1 unit/ml) in 100 mM triethanolamine-HCL, 6.5 mM MgCl_2 , 2.7 mM Na_2ATP and 0.83 mM NaNAD at pH 8 at 30°C. All following steps were as above.

The enzymatic assay mixtures for the generation of MAL, OAA, α KG were measured using the organic acid LC-MS/MS method from section 7.5. The measured mass transitions of each metabolite were acquired with settings of CE, DP and cell exit potential identified in a tuning with unlabeled analytical standard of the metabolite in question if not otherwise stated. Both mass transitions and CE, DP and CXP settings can be found in Electronic Supplementary Material S3 of Kappelmann et al. [91]. The sugar phosphates 6PG, F6P, 6PG and FBP were measured using the sugar phosphate method from Table 3.8a. Both mass transitions and MS parameter settings can be found in Electronic Supplementary Material S3 of Kappelmann et al. [91]. CAD pressure was 5.

Selectively labeled CIT was generated by using a modified version of the Pyruvate carboxylase assay from above. Malate dehydrogenase and NADH were omitted from the assay solution. Substrates were sodium [1- 13 C]pyruvate, sodium [2- 13 C]pyruvate and sodium [3- 13 C]pyruvate in combination with either NaHCO_3 or $\text{NaH}[^{13}\text{C}]\text{O}_3$ and unlabeled ACCOA. Additionally, the assay solution contained 0.1 mM 5,5'-dithiobis(2-nitrobenzoic acid) (dissolved in absolute ethanol at a concentration of 10 mM) to check the success of the enzymatic reaction photometrically. Before use in the assay, citrate synthase was diluted 1:10 in 3.2 M $(\text{NH}_4)_2\text{SO}_4$ at pH 7. From this solution 3 μ l was added to the assay, yielding a final enzyme concentration of 1 unit/ml. The reaction was again started by adding 3 μ l Pyruvate carboxylase, followed by gentle agitation. After 15 minutes of incubation at 30°C, the samples were measured at 412 nm, processed and stored as described above. For LC-ESI-MS/MS analysis the assay samples were diluted with acetonitrile to a concentration of 60 % (v/v) and stored at -80°C till analysis. The same protocol was used for the production of ^{13}C -isotopic standards of ICIT and cis-ACO, except that the assay solution was at pH 7.4. Additionally, 0.05 units/ml of Aconitase, which was activated according to the manufacturer protocol [166], were added to the assay. For the production of [*hydroxy*- ^{18}O]CIT and [*hydroxy*- ^{18}O]ICIT 250 μ l of a 25 mM solution of cis-ACO in 100 mM Tris-HCL at pH 7.4 were evaporated to dryness in a vacuum centrifuge. After the residue was dissolved in the same volume of $\text{H}_2[^{18}\text{O}]$, 0.05 units/ml Aconitase were added. All subsequent steps were carried out as described above.

The enzymatic assays to generate selectively labeled CIT, ICIT and cis-ACO were separated using the HPLC column described in the study of Teleki et al. [170] with the following eluent profile at a flowrate of 100 μ l min^{-1} : 0 min: 10% A, 2 min: 10% A, 26 min: 38% A, 50 min: 58% A, 50.05 min: 90% A, 60 min: 90% A, 80 min: 10% A, followed by 15 min equilibration at 10% A at a flow of 200 μ l/min. Buffers were A: 10 mM $\text{NH}_4\text{CH}_3\text{COO}$, pH 9.2; B: acetonitrile. The HPLC outlet was coupled to the ESI

interface of the QqTOF MS. During the elution the Q1 windows corresponding to mass isotopomers of CIT (CE: -20 eV, CE spread: 10 eV and DP: -15 eV) and cis-ACO (CE: -15 eV, CE spread: 10 eV and DP: -60 eV) were monitored. The same HPLC method was used to separate the assays for production of [*hydroxy*-¹⁸O]-isotopes of CIT and ICIT.

7.10 Estimation of extracellular rates

Many studies employ ordinary difference quotients to estimate extracellular rates. Following this method, the specific substrate uptake rate is estimated to be:

$$\Pi_{\text{Glc, spe}}(t_i) = \frac{c_{\text{Glc}}(t_{i+1}) - c_{\text{Glc}}(t_i)}{(X(t_{i+1}) - X(t_i))(t_{i+1} - t_i)} \quad (3.3)$$

In equation (3.3) $c_{\text{Glc}}(t_i)$ denotes the glucose concentration in mmol l^{-1} at time point t_i , whereas $X(t_i)$ is the biomass concentration at this time point. The difference quotient method has two main drawbacks: As linear approximation to the true exponential growth curve, it becomes ever less accurate towards the end of the growth phase. Furthermore, the expression in equation (3.3) derives the uptake rate from only two measurements for both glucose and biomass concentration.

Therefore, a model based approach, reconciling all available measurement data at once was applied in this work. Using this approach the estimation of the specific growth rate μ and specific substrate uptake rates $\Pi_{\text{Glc, spec}}$ is accomplished by maximum likelihood estimation of parameters of an algebraic expression describing the biomass growth and concurrent substrate utilization during the exponential phase of a batch fermentation. In the following this expression shall be derived. The definitions and units of all used quantities can be found in the Nomenclature section of this thesis.

The remaining substrate at any given time point is the integral over the volumetric uptake rate in mmol (l h)^{-1} :

$$c_{\text{Glc}}(t) = c_{\text{Glc},0} - \int_{t_0}^t \Pi_{\text{Glc, vol}}(\hat{t}) d\hat{t} \quad (3.4)$$

Assuming a constant specific substrate uptake the volumetric glucose uptake rate can be expressed in the following way:

$$\Pi_{\text{Glc, vol}}(t) = \Pi_{\text{Glc, spe}} X(t) \quad (3.5)$$

where $X(t)$ denotes the biomass concentration in $\mu\text{l ml}^{-1}$ or g l^{-1} , depending on whether the biomass signal is BV or CDW. In the latter case the unit of $\Pi_{\text{Glc, spe}}$ is $\text{mmol (g CDW h)}^{-1}$, in the former case $\text{mmol } (\mu\text{l BV h)}^{-1}$. The increase in biomass concentration during the exponential phase can be expressed as:

$$X(t) = X_0 \exp(\mu t) \quad (3.6)$$

Inserting equation (3.5) and (3.6) into the integral of equation (3.4) and carrying out the integra-

tion yields:

$$c_{\text{Glc}}(t) = c_0 - \frac{\Pi_{\text{Glc, spe}}}{\mu} X_0 \exp(\mu t) \quad (3.7)$$

Given this set of equations, the rate estimation is based on a variance-weighted least-squares fit, fitting equations (3.6) and (3.7) jointly to the experimentally observed time courses of biomass and substrate concentration, respectively. The parameters are $\Pi_{\text{Glc, spe}}$, X_0 , μ , and c_0 , where c_0 absorbs the integration constant and the initial substrate concentration. Only data points within the exponential phase were included into the fit. The end of the exponential phase was judged by the peak in CO_2 volume fraction in the exhaust gas stream.

The fitting procedure is carried out using a sequential quadratic programming optimization routine from MATLAB [118]. The estimation of confidence intervals is based on a parametric Monte Carlo bootstrapping approach, analogous to the bootstrap method implemented in the 13CFLUX2 package and described in Dalman et al. [39]. Essentially, every available measurement is perturbed independently within the normal distribution described by the measurement value and its standard deviation of the respective measurement. The perturbation operation consists in sampling from the normal distribution of each measurement point and inserting the sample as measurements. The above described variance-weighted least-square fit is then re-performed. The sample was generated as a Latin-Hypercube sample using the *lhs*-function of MATLAB and comprises 1000 samples if not otherwise stated. The upper and lower confidence bounds are then given as the α -th and $(1-\alpha)$ -th percentile of the parameter sample obtained from 1000 least-squared fits. If not otherwise stated α is set to 0.25, the confidence interval being the Interquartile Range (IQR).

The total CO_2 -formation rate is calculated by equation (3.8), which arises from balancing the gas-phase of the bioreactor. In this equation Φ denotes the volume fraction of the gas species in question in vol%, whereas the superscripts α and ω denote inlet and outlet concentrations, respectively. When the inlet air flow F_{air} is in $\text{m}^3 \text{h}^{-1}$, the calculated total CO_2 -formation rate is in mol h^{-1} .

$$\hat{\Pi}_{\text{CO}_2, \text{tot}}(t) = \frac{F_{\text{air}} P_{\text{ges}}}{RT} \left[\frac{(100 - \Phi_{\text{O}_2}^{\alpha} - \Phi_{\text{CO}_2}^{\alpha})}{(100 - \Phi_{\text{O}_2}^{\omega}(t) - \Phi_{\text{CO}_2}^{\omega}(t))} * \frac{\Phi_{\text{CO}_2}^{\omega}(t)}{100} - \frac{\Phi_{\text{CO}_2}^{\alpha}}{100} \right] \quad (3.8)$$

To obtain the biomass specific carbon dioxide formation rate $\Pi_{\text{CO}_2, \text{spec}}(t)$, the total rate $\Pi_{\text{CO}_2, \text{tot}}(t)$ was divided by the total biomass at a time point t . To compute the latter one from the biomass concentration measurements, the liquid volume in the bioreactor at any given time needs to be known. This was achieved by accurately tracking the sample volume withdrawn for analytics during the course of cultivation.

7.11 Full information FIA

In general, identifiability relies on the dimension of the solution space for both $\mathbf{x}$ and $\mathbf{v}_f$ of equation (2.19) for a given $\mathbf{x}_{\text{inp}}$. This can in principle be decided based on the Gröbner base algorithm as shown by Wiechert [192]. However, due to the computational complexity of the problem, Isermann

and Wiechert [81] introduced a very powerful FIA method which is structural, global, exact, as well as unique. The only drawback is that it assumes full measurement information, i.e. $\mathbf{g}(\mathbf{x}) = \mathbf{x}$. The algorithm to compute rDOF proceeds in the following steps:

1. Randomly choose rational vectors $\mathbf{v}_f$ of feasible intracellular free fluxes and input labels $\mathbf{x}_{\text{inp}}$. Here rational means that all numbers are represented as exact rational numbers.
2. Calculate the resulting labeling distribution $\mathbf{x}(\mathbf{v}_f, \mathbf{x}_{\text{inp}})$ by solving equations using rational number arithmetics. The forward problem, i.e. the simulation of the cumomer fraction vector for a given free flux vector, can be solved by computing a cascade of linear systems as shown by Wiechert et al. [195] and is outlined in the Supplementary Information of Kappelmann et al. [90]. The result contains only (very large but exact) rational numbers as follows from the solution theory of labeling systems.
3. Now assume the complete calculated data set $\mathbf{x}(\mathbf{v}_f, \mathbf{x}_{\text{inp}})$ as the measured data and substitute it back into the original equation system (2.19). Since $\mathbf{x}$ is now given by numbers, this equation turns out to be linear w.r.t the free fluxes $\mathbf{v}_f$. It can thus be written as a linear equation system:

$$\mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp}}) \cdot \mathbf{v}_f = \mathbf{m}(\mathbf{x}, \mathbf{x}_{\text{inp}}) \quad (3.9)$$

with the known flux information matrix $\mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp}})$ and r.h.s. $\mathbf{m}(\mathbf{x}, \mathbf{x}_{\text{inp}})$ [81].

4. Since $\mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp}})$ is a known matrix of rational numbers, its rank can now be exactly determined. A unique solution for $\mathbf{v}_f$ can be derived from the linear system when the rDOF are zero:

$$\text{rDOF} = \dim(\mathbf{v}_f) - \text{rank}(\mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp}})) = 0 \quad (3.10)$$

The solution is unique since a full-rank linear system has only one unique solution. Remarkably, it can be shown that the result of the rank calculation is independent of the chosen $\mathbf{v}_f$ and $\mathbf{x}_{\text{inp}}$ (i.e. global) if they are chosen 'sufficiently general'. This means that a random number generator will suggest $\mathbf{v}_f$ and $\mathbf{x}_{\text{inp}}$ which almost surely (i.e. with probability one) produce the correct matrix rank. For a detailed example the reader is referred to Supplementary Information S5 of [90]. The random number generator can also be substituted by repeatedly performing the above algorithm and deliberate variation of the chosen substrate compositions.

Results and Discussion - Part I

8 Structural Identifiability Analysis of anaplerotic reactions by ^{13}C -MFA

8.1 Motivation

The anaplerotic node in *C. glutamicum* has been elusive so far by ^{13}C -MFA [179, 134] since inflated confidence intervals were reproducibly obtained for the reactions at this metabolic branch point. The *a posteriori* non-identifiable reactions may be indicative of either statistical or structural non-identifiability, both of which reflect themselves in the same observed inflated *a posteriori* confidence bounds. However, the reason for this observation needs to be established as drastically different consequences arise for each underlying cause. If statistical identifiability lies to the core of the problem, more comprehensive and more precise measurements, possibly integrated over several experiments will be required. If the problem is structural in nature, the model needs to be adapted accordingly since the reactions would remain non-identifiable regardless of the experimental effort. Therefore, the following section is devoted to establishing the nature of the non-identifiability of the anaplerotic reactions in *C. glutamicum*.

8.2 Network reduction for exact flux identifiability analysis

The *a posteriori* non-identifiable anaplerotic reactions in the above mentioned studies were obtained using a comprehensive model of *C. glutamicum*. Such a model is depicted in Figure 4.1. This model is adapted from van Ooyen et al. [179] to include the CO_2 -dependent biosynthesis reactions of inosine 5-monophosphate (IMP), representing purine biosynthesis and uridine 5-monophosphate (UMP), representing pyrimidine nucleotide biosynthesis. By encompassing these reactions, the model comprises all reactions involving net production, consumption and label exchange of CO_2 under conditions optimizing biomass growth, and thereby enables the usage of the measured CO_2 label fraction in CO_2 for flux inference. The biosynthesis of fatty acids as well as mycolic acids fix CO_2 but the very same CO_2 molecule is released at a subsequent enzymatic step such that these reactions neither contribute to net production nor to the labeling equilibrium.

To constrain the biomass withdrawal reactions the biomass composition of Marx et al. [117] has been used in many studies so far [201, 134, 179, 197, 163]. However, except for the amino acid content the compositional information is entirely taken from an *E. coli* data set [129]. It includes lipopolysaccharides which play only a marginal role in *C. glutamicum* [46]. Furthermore, it does not include mycolic acids in biomass which were estimated to account for approximately 10 % of the CDW in *C. glutamicum* [150, 15]. For the present study, instead, the compositional information provided in Kjeldsen and Nielsen [96] was used. In the authors' opinion it represents the most recent and most comprehensive compilation of compositional data on *C. glutamicum* biomass available. Since all amino acids as well as pyrimidine and purine nucleotides are modeled as metabolite pools the num-

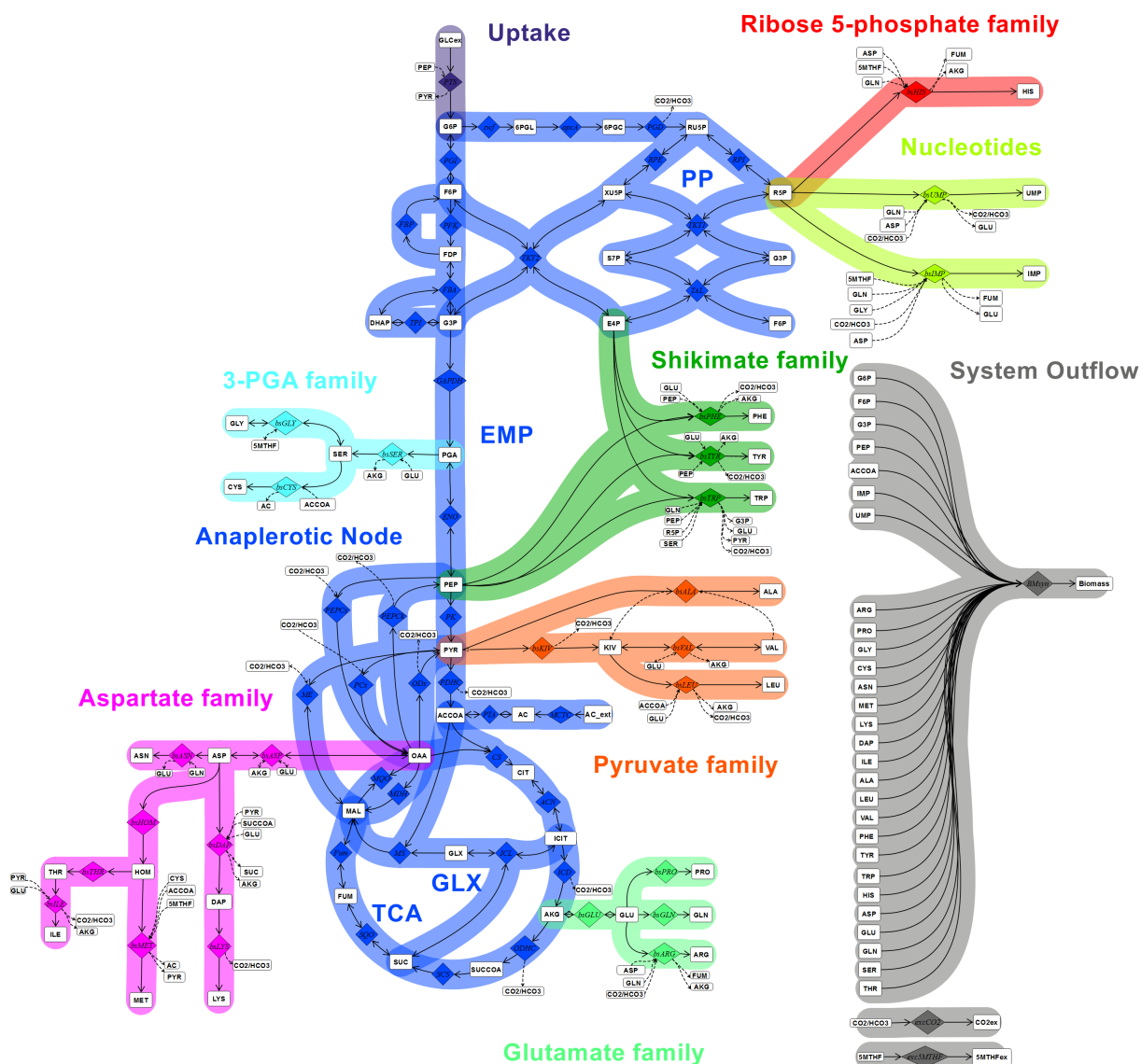


Figure 4.1: Comprehensive model of *C. glutamicum* central carbon metabolism encompassing glycolysis, anaplerotic reactions, tricarboxylic cycle and biosynthesis reactions for amino acids as well as the purine nucleotide IMP and the pyrimidine nucleotide UMP. For all balanced metabolites in this network specific precursor demand was computed from the biomass composition given in Kjeldsen and Nielsen [96], assuming standard stoichiometric relationships between modeled metabolites and biomass polymers.

ber of central metabolism "precursor metabolites" reduces from eleven in Marx et al. [117] to five, to wit G6P, F6P, GAP, PEP and ACCOA. The remaining biomass precursor pools are amino acids, IMP and UMP. The content of biomass constituents on a CDW content basis are all taken from Kjeldsen and Nielsen [96]. To compute the efflux from biomass precursors the stoichiometry of biosynthesis reactions was taken from Neidhardt et al. [129] and the one for Corynemycolates from Eggeling and Bott [46].

The structural flux identifiability problem under isotopically stationary conditions in general is concerned with finding a solution for $\mathbf{v}_f$ from equation (2.19) given the vector of measured labeling enrichments $\mathbf{y}$. This can be answered by analyzing the rank of a method-dependent matrix. However, the model of Figure 4.1 comprises far too many metabolite pools to enable an exact rank estimation by means of integer arithmetics.

To analyze the identifiability for different enzymatic activities at the anaplerotic node in *C. glutamicum* and to enable exact computations and accurate rank estimation, the focused network of Figure 4.2 was considered, which represents a valid reduction of the comprehensive network model to the anaplerotic node. As a starting point, this network comprises all enzymatic activities at the anaplerotic node as annotated in *C. glutamicum* ATCC 13032 [88]. Blue dashed lines denote reactions of which the *in vivo* activity is negligible under all conditions considered here, based on biochemical evidence (see Supplementary Information of Kappelmann et al. [90]). Blue dashed lines may also represent co-factor usage which is not considered for ^{13}C -isotope balancing. Black solid lines represent reactions which are relevant under solely glycolytic conditions and are implemented as carbon transitions for the simplified isotopomer model describing the enzymatic set of *C. glutamicum* wild type grown on glucose.

It is a well established fact that *C. glutamicum* is able to metabolize a broad spectrum of organic acids and glucose simultaneously with constant specific uptake rates [140, 189]. The problem of identifiability is therefore also considered for the co-metabolization of lactate (LAC) and acetate. However, acetate metabolism is known to induce the expression of *aceA* and *aceB* in *C. glutamicum*, giving rise to two new carbon atom transitions catalyzed by ICL and MS [57]. These carbon atom transitions, together with the corresponding transitions underlying the assimilation of both organic acids are additionally considered when analyzing the anaplerotic node during co-metabolism.

The box surrounding the metabolites PEP, PYR, OAA and MAL depicts the system boundary and defines the 3-phosphoglycerate (PGA), LAC, acetate, CO_2 and ASP pool as external pools. The isotopomer distribution of PGA can be assumed to be known from the defined mixture of labeled glucose as well as the measured flux distribution in the upper glycolysis and PP pathway (see Supplementary Information of Kappelmann et al. [90]). The TCA cycle is modeled as one overall reaction sequence with two different atom transitions to account for the label scramble at the SUC/FUM pool conforming with the network reduction rules outlined in van Winden et al. [180]. In the metabolic network at hand the system's influxes (ν_{ENO} , ν_{PTA} , ν_{LHD}) as well as outflux (ν_{ASAT}) are assumed to be measured. The latter one can be assumed to arise from employing a biomass formation equation and

additional product formation rates for TCA cycle derived products, either of which provides a value for the TCA cycle efflux. Further justifications for the employed network reductions are given in the Supplementary Information of Kappelmann et al. [90].

As apparent from Figure 4.2, the CO_2 pool in the focused network model represents the only critical system influx. The comprehensive model from above models all CO_2 -dependent reactions and thus allows to compute the labeling fraction in CO_2 during forward simulation. By contrast, this focused network considers CO_2 as external pool, being supplied as measurement. This can be reliably performed by infrared spectroscopy of the exhaust gas stream [50] or membrane inlet mass spectrometry [201]. Externally supplied ^{13}C may be considered as substrate itself for specific questions [116]. However, when ^{13}C -labeled glucose is used as substrate the CO_2 label distribution cannot be deliberately chosen but arises from the chosen substrate isotopomer composition, carboxylation and decarboxylation activity of the network as well as experimental parameters [108]. Regardless of the experimental setup or organism in question the fractional enrichment will result in a value between 0 and 1. For the focused network model it turns out that the rank of $\mathbf{M}$ is maximal as long as the CO_2 labeling fraction is other than 0 and 1, and the substrate composition of the PGA pool does not entirely consist of the fully labeled or unlabeled species. One mixture fulfilling this criterion was used for all simulations (see Supplementary Information of Kappelmann et al. [90]) and a CO_2 label of 0.2 was assumed. Although the CO_2 will most probably assume a different value under real experimental conditions the decision on the maximal rank of $\mathbf{M}$ is unaffected by its actual value as long as it is other than 0 or 1.

The simplified isotopomer model was implemented in Mathematica [199] and all computations under isotopically stationary conditions were performed in the space of rational numbers. The implementation is provided as Mathematica notebook as Supplementary Information of Kappelmann et al. [90]. For each scenario of anaplerotic activity, the network was changed accordingly, the corresponding flux variable removed from the flux vector and the network simulated for one non-trivial setting of free fluxes.

8.3 Full information FIA in a focused network model

For each network variant the rank analysis of matrix $\mathbf{M}$ was performed according to section 7.11, given the simulated cumomer vector supplied as exact numbers. This analysis enables to decide unambiguously and without any numerical artifacts whether all fluxes of a certain labeling network can be inferred at all from labeling measurements. This decision can be made since the scenario of full information is considered. Knowing all system influxes and outfluxes as well as the entire labeling state vector represents the scenario of maximal possible information. With regard to equation (2.19) $\mathbf{g}(\mathbf{x})=\mathbf{x}$ holds. The results of this analysis for different network topologies and substrate compositions are presented in Table 4.1. Therein the omission of a certain reaction and removal of the corresponding carbon transition is denoted by its enzyme name.

It can be seen that the full set of enzymatic activities describing *C. glutamicum* wild type on glu-

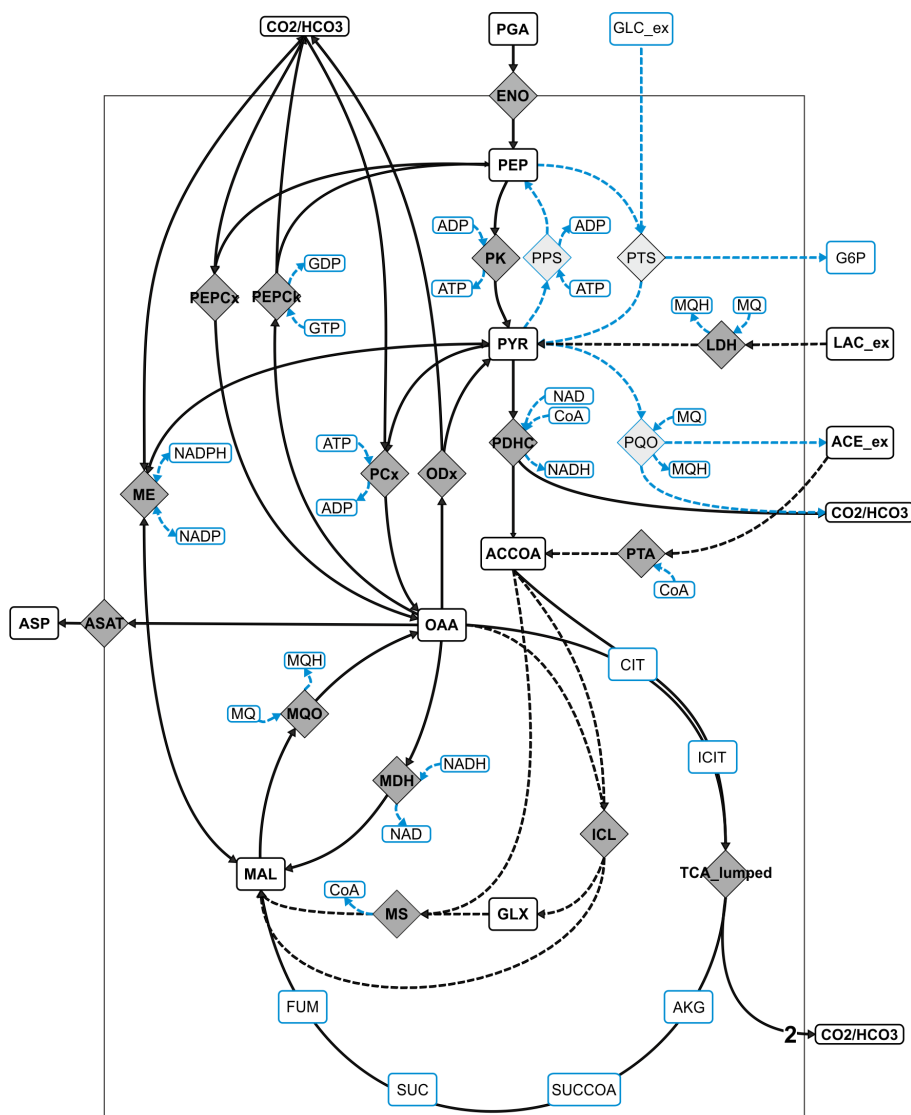


Figure 4.2: Focused isotopomer model of the anaplerotic node in *C. glutamicum*. The transparent blue-framed metabolites were not balanced. The network topology describing growth on glucose comprises only the reactions represented by solid black lines. Black dashed lines represent additional carbon transitions considered for cometabolization of glucose and lactate or acetate. Reactions and metabolites colored in blue represent annotated but negligible reactions due to biochemical evidence or network simplification without loss of information. Further details may be found in the Supplementary Information of Kappelmann et al. [90]. Abbreviations: ASAT, Aspartate transaminase; LDH, Lactate dehydrogenase; ENO, Enolase; ICL, Isocitrate lyase; MDH, Malate dehydrogenase; ME, Malic enzyme; MQO, Malate:quinone oxidoreductase; MS, Malate synthase; ODx, Oxaloacetate decarboxylase; PCx, Pyruvate carboxylase; PDHC, Pyruvate dehydrogenase complex; PEPCx, Phosphoenolpyruvate carboxylase; PPS, Phosphoenolpyruvate synthetase; PQO, Pyruvate:quinone oxidoreductase; PTA, lumped reaction of Acetate kinase and Phosphotransacetylase; PTS, Phosphotransferase system.

cose still leaves one rDOF, which means that the value of one unknown flux must be constrained for the remaining fluxes to be identifiable. Given this result it can already be concluded that the high standard deviations on anaplerotic fluxes in the studies of van Ooyen et al. [179] and Noack et al.

Table 4.1: Identifiability of anaplerotic reactions in *C. glutamicum* under full information. The rDOF for different network topologies and substrate compositions were computed by integer arithmetic applying equation (3.10). The topologies for all combinations of single and double omissions of enzymatic activities of PEPCx, PCx, PEPCK, ODx and ME are compared to the wild-type (WT) network comprising the full set of anaplerotic reactions. The deletion mutant in ME and MDH corresponds to the network topology used in the study of Petersen et al. [144] with lumping of the OAA and MAL pool.

Network topology	rDOF			
	Glucose	Glucose Lactate	Glucose Acetate	Glucose Lactate Acetate
WT	1	1	1	1
single enzymatic omissions				
ΔODx	0	0	0	0
ΔPCx	1	1	1	1
ΔPEPCx	2	2	2	2
ΔPEPCK	1	1	1	1
ΔME	0	0	0	0
double enzymatic omissions				
$\Delta(\text{PEPCx ME})$	0	0	0	0
$\Delta(\text{PCx ME})$	0	0	0	0
$\Delta(\text{PCx PEPCx})$	2	2	2	2
$\Delta(\text{PEPCK PCx})$	1	1	1	1
$\Delta(\text{PEPCK PEPCx})$	2	2	2	2
$\Delta(\text{PEPCx ODx})$	1	1	1	1
$\Delta(\text{ODx ME})$	1	0	1	0
$\Delta(\text{ODx PCx})$	0	0	0	0
$\Delta(\text{PEPCK ME})$	0	0	0	0
$\Delta(\text{PEPCK ODx})$	0	0	0	0
$\Delta(\text{ME MDH})$	0	0	0	0

[134] are structural in nature and not a purely statistical result.

Since rDOF gives the number of assumptions on flux directionality or additional non-redundant flux measurements necessary to determine all fluxes, a network must be considered where that allows to make such assumptions. A gene-knockout of the corresponding reaction can be quickly realized by the omission of the reaction, thereby reducing the number of fluxes to be estimated from the same set of label information. As can be seen in Table 4.1, the removal of one reaction leads to the counter-intuitive result that the resulting rDOF are not zero, but may even increase as evidenced by the ΔPEPCx strain. This means that by changing the number of variables (the dimension of ν_f), the rank of matrix $\mathbf{M}$ may also change in equation (3.10) which makes the change of rDOF upon removal of a reaction *a priori* unpredictable. Missing reactions alter the network structure and induce a different ^{13}C -label flow compared to the initially considered network, reducing the information content of the generated label distribution. Therefore a detailed analysis is needed accounting for the altered equilibrium ^{13}C -label fractions in an altered isotopomer network.

The increase of rDOF for the Δ PEPCx strain still holds true for the double enzymatic omissions comprising PEPCx and ODx. Here it is worth noting that the presented analysis is directly transferable to other compartment-free organisms. For example, the latter combination also represents the enzymatic activity at the anaplerotic node of *Bacillus subtilis* with PCx as sole anaplerotic enzyme and a missing ODx activity [158]. Thus, the rank deficiency of one for this network topology may only be circumvented by assuming ME to act unidirectionally. By contrast, *Escherichia coli* lacks PCx as well as ODx which, following the above analysis, permits the unambiguous flux determination at the anaplerotic node.

Intuitively, fluxes impinging on one metabolite pool can be differentiated if the isotopomer distributions of the influxes are different either through a different isotopomer mapping matrix describing the influx or through a differently labeled influx pool in the case of identical isotopomer mappings. Within this fact lies the clue to experimental design and the motivation for the usage of co-substrates. For the measurement of the carboxylation reactions Petersen et al. [144] employed LAC as co-substrate to enforce a differentially labeled PEP and PYR pool since the isotopomer mapping of the anaplerotic carboxylation and decarboxylation reactions is identical. However, when ME and ODx are active, they will inherently proportion a differential labeling between PEP and PYR enabling the carboxylation reactions to be distinguishable (cf. Figure 4.2). This explains why LAC as co-substrate decreases the rDOF only if ME and ODx are both inactive (see Table 4.1). By contrast, acetate as co-substrate does not provide an additional differential labeling of ACCOA which might not also be generated under solely glycolytic conditions.

The above result also highlights a general limitation of the running identifiability analysis. It is only valid for *a priori* conclusions on flux identifiability if all reactions assumed to be present are indeed active and induce a label flow. For real metabolic networks this decision is not trivial, especially when *in vitro* enzymatic assays suggest activity which may not be present under *in vivo* conditions due to inhibitory regulation. Second, it does not account for parameter sensitivities below measurement noise ('statistical identifiability'). Measurement noise may restrict the significance of this result, for example when the magnitude of the exchange flux between MAL and OAA becomes large. In the limiting case of a very large exchange flux the labeling of both pools is equal, thus impeding a differentiation between ODx and ME.

The rank of matrix $\mathbf{M}$ for an arbitrary network can maximally be $\dim(\mathbf{v}_f) - 1$ when the system influx is not known [81]. Thus, the result of an identifiability analysis by ^{13}C -tracers inherently depends on the assumptions on which fluxes are already determined by rate measurements. In compartment-free organisms the system influx is usually equivalent to the substrate uptake rate which is amenable to experimental determination. However, this may not be the case for fluxes over system boundaries present in networks describing for example mammalian cells [131] or host-pathogen interactions [20].

Various studies uncovered the advantage of parallel labeling experiments for improved precision of flux inference [109] and increased number of estimable fluxes [38]. However, the latter claim has

been made based on a statistical criterion, i.e. confidence intervals derived from real experimental data. In terms of a structural analysis parallel labeling experiments correspond to solving two linear equation systems of type (3.9) for two different input cumomer vectors $\mathbf{x}_{\text{inp}}$ simultaneously according to:

$$\begin{pmatrix} \mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp},1}) \\ \mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp},2}) \end{pmatrix} \cdot \mathbf{v}_f = \begin{pmatrix} \mathbf{m}(\mathbf{x}, \mathbf{x}_{\text{inp},1}) \\ \mathbf{m}(\mathbf{x}, \mathbf{x}_{\text{inp},2}) \end{pmatrix} \quad (4.1)$$

Again, a unique solution can be obtained if the concatenated matrix has full rank. Each $\mathbf{x}$ in $\mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp}})$ is dependent upon $\mathbf{x}_{\text{inp}}$ and $\mathbf{v}_f$. For given values of $\mathbf{v}_f$ the dependence $\mathbf{x}(\mathbf{x}_{\text{inp}}, \mathbf{v}_f)$ can be analytically computed and substituted for $\mathbf{x}$, yielding $\mathbf{M}(\mathbf{v}_f, \mathbf{x}_{\text{inp}})$. By providing $\mathbf{x}_{\text{inp},1}$ and $\mathbf{x}_{\text{inp},2}$ as variables rather than numerical values the rank estimation function in Mathematica automatically assumes both variable vectors to be independent. Thus, if an additional tracer were to increase the structural information gain additional linearly independent equations would be included in the rank estimation through $\mathbf{M}(\mathbf{x}, \mathbf{x}_{\text{inp},2})$ in equation (4.1), resulting in a higher -or optimally- full rank and thus improved structural information.

Performing a rank analysis of the concatenation of two matrices $\mathbf{M}$ for the *C. glutamicum* WT network topology with two different sets of variables for $\mathbf{x}_{\text{inp},1}$ and $\mathbf{x}_{\text{inp},2}$ reveals that the maximal rank is not increased compared to an evaluation on a single tracer. Thus, an additional tracer does only yield linear dependent equations. With the argument of section 7.11 this holds globally for all $\mathbf{v}_f$. This has a consequence for any further parallel experiment included in the joint inference: As proven in the Supplementary Information S6 of Kappelmann et al. [90] no further parallel experiment leads to a structural information gain when already two experiments do not yield a rank increase. A statistical information gain based on confidence intervals, however, has already several times been shown to be increased by the use of parallel tracers [37, 38].

8.4 Partial information FIA in a focused network model

In any real experimental setup only linear combinations of isotopomer fractions are available as measurements. This limitation may further restrict the inference of fluxes. If only a part of the labeling state vector $\mathbf{x}$ is measured, $\mathbf{g}(\mathbf{x})$ is not the identity map. In that case Isermann and Wiechert [81] and van Winden et al. [180] propose to analyze the rank of the Jacobian of the measured isotopomer fractions $\mathbf{y}$ with respect to the free fluxes $\mathbf{v}_f$. These derivatives are calculated according to the chain rule:

$$\frac{\partial \mathbf{y}(\mathbf{x}_{\text{inp}}, \mathbf{v}_f)}{\partial \mathbf{v}_f} = \frac{\partial \mathbf{y}(\mathbf{x})}{\partial \mathbf{x}} \cdot \frac{\partial \mathbf{x}(\mathbf{x}_{\text{inp}}, \mathbf{v}_f)}{\partial \mathbf{v}_f} \quad (4.2)$$

$\mathbf{x}$ denotes the labeling state vector for each metabolite considered in the measurement system. The derivative $\frac{\partial \mathbf{x}}{\partial \mathbf{v}_f}$ is calculated by implicit differentiation of equation (2.19) as outlined in the Supplementary Information of Kappelmann et al. [90] according to Wiechert et al. [195]. Based on the implicit function theorem, this method only yields locally unique flux estimators (see section 5.3.3).

Experimentally, part of the isotopomer distribution is accessible by a variety of NMR [23], GC-MS [18] and LC-MS/MS [156] techniques, expressed by the function $\mathbf{g}(\mathbf{x})$. In the following we focus on the mass transitions on a LC-MS/MS device in multiple reaction monitoring mode because this technique proved most suitable for central metabolism intermediates as those at the anaplerotic node [200, 114]. Table S3 in the Supplementary Information of Kappelmann et al. [90] lists the currently observable mass transitions for the measurement setup of the running analysis. For a partial information FIA of the focused network model the Jacobian matrix of equation (4.2) was evaluated using exact integer arithmetic. As a starting point, the measurement of the mass transitions of all four anaplerotic metabolites Mal, OAA, PYR and PEP given in Table S3 of the Supplementary Information of Kappelmann et al. [90] was simulated. However, metabolites PYR and OAA represent unstable α -ketoacids of which the isotopomer distribution may not be measurable due to degradation [209, 164]. Therefore different measurement models were considered comprising various combinations of the mass transitions of the four anaplerotic metabolites.

The focused model of section 8.2 was analyzed with respect to partial information using exact integer arithmetic to enable exact rank computations of the Jacobian under all circumstances. The results of this analysis in Table 4.2 show that the restricted knowledge of only experimentally measurable mass transitions suffices to attain the same level of flux information as in the case of full information (Table 4.1). At first sight this may seem counter intuitive but resides in the fact that many labeling balances are redundant. For instance, the OAA cumomers x_{1xxx}^{OAA} , x_{x1xx}^{OAA} , and x_{xx1x}^{OAA} all contain the same information with respect to the carboxylation fluxes since no change in the carbon skeleton occurs between the three indicated positions. Each of these cumomer fractions behaves identically with respect to a change of a flux in the system.

Recruiting the label information of further metabolites will prove elusive. The TCA cycle can be described as one sequence of linear mappings of isotopomer distributions from CIT to MAL. Therefore, measuring label enrichment in TCA cycle intermediates upstream to MAL does not provide additional information which is not already contained in the carbon backbone of MAL. The same holds true for the amino acids ASP and ALA the isotopomer distribution of which is linearly derived from the ones of OAA and PYR respectively. Consequently, their inclusion in the rank estimation will only add linearly dependent equations.

Table 4.2 also reveals that the information content of single metabolites varies considerably with the network architecture. A striking example is the single measurement of mass transitions of PEP in the Δ (PEPCK ME) network topology. For this topology the number of variables equals the rDOF such that no information is gained from measuring the mass isotopomer distribution of PEP. This is obvious since in the absence of an influx into this metabolite pool (catalyzed by PEPCK) its label is identical to the known substrate label distribution of the PGA pool (cf. Figure 4.2), thus it is completely uninformative.

Although LAC as co-substrate does not reduce the rDOF in the case of full information, it does so when only one single or two metabolites are measured (cf. Table 4.2). For instance, knowledge of

Table 4.2: Identifiability of anaplerotic reactions in *C. glutamicum* under partial information with either glucose as sole carbon substrate or during glucose and lactate co-metabolization. For each measurement scenario and network topology the rDOF are determined applying equation (4.4). The terms $\mathbf{y}_{\text{PEP}}$, $\mathbf{y}_{\text{PYR}}$, $\mathbf{y}_{\text{OAA}}$, $\mathbf{y}_{\text{MAL}}$ denote the inclusion of the corresponding mass transitions into the measurement system according to Table S3 of the Supplementary Information of Kappelmann et al. [90].

		rDOF						
		WT	$\Delta(\text{PCx ODx})$	$\Delta(\text{PEPCx ME})$	$\Delta(\text{PEPCk ME})$	ΔME	ΔPEPCx	ΔPCx
Glucose	$\dim(\mathbf{v}_f)$	6	4	3	3	4	5	5
	full information	1	0	0	0	0	2	1
	$\mathbf{y}_{\text{PEP}}, \mathbf{y}_{\text{PYR}}, \mathbf{y}_{\text{OAA}}, \mathbf{y}_{\text{MAL}}$	1	0	0	0	0	2	1
	$\mathbf{y}_{\text{OAA}}$	4	2	2	1	2	4	3
	$\mathbf{y}_{\text{MAL}}$	3	1	1	0	1	3	2
	$\mathbf{y}_{\text{PEP}}$	4	2	1	3	2	3	3
	$\mathbf{y}_{\text{PYR}}$	5	3	2	2	3	4	4
	$\mathbf{y}_{\text{MAL}}, \mathbf{y}_{\text{OAA}}$	2	0	1	0	1	3	1
	$\mathbf{y}_{\text{OAA}}, \mathbf{y}_{\text{PYR}}$	4	2	2	1	2	4	3
	$\mathbf{y}_{\text{MAL}}, \mathbf{y}_{\text{PEP}}$	2	0	0	0	0	2	1
Glucose Lactate	$\dim(\mathbf{v}_f)$	6	4	3	3	4	5	5
	full information	1	0	0	0	0	2	1
	$\mathbf{y}_{\text{PEP}}, \mathbf{y}_{\text{PYR}}, \mathbf{y}_{\text{OAA}}, \mathbf{y}_{\text{MAL}}$	1	0	0	0	0	2	1
	$\mathbf{y}_{\text{OAA}}$	3	1	2	1	1	4	2
	$\mathbf{y}_{\text{MAL}}$	2	0	1	0	0	3	1
	$\mathbf{y}_{\text{PEP}}$	3	1	1	3	1	3	2
	$\mathbf{y}_{\text{PYR}}$	4	2	2	1	2	4	3
	$\mathbf{y}_{\text{MAL}}, \mathbf{y}_{\text{OAA}}$	1	0	1	0	0	3	1
	$\mathbf{y}_{\text{OAA}}, \mathbf{y}_{\text{PYR}}$	3	1	2	1	1	4	2
	$\mathbf{y}_{\text{MAL}}, \mathbf{y}_{\text{PEP}}$	1	0	0	0	0	2	0

the mass isotopomer distribution of MAL suffices to completely identify the network topology of the $\Delta(\text{PCx ODx})$ strain under glucose and LAC, whereas this measurement does not guarantee $\text{rDOF} = 0$ under solely glycolytic conditions. This effect vanishes when the isotopomer distribution of LAC is equal to the one of the PGA pool. Therefore, co-substrates, as long as their label is differential to the primary substrate, may structurally increase the information content of single metabolites.

8.5 Non-unique global FIA in a comprehensive isotopomer model of *C. glutamicum*

Certainly, using a reduced model for testing flux identifiability lets arise the question of whether it is a valid representation of a comprehensive network with practical relevance for ^{13}C -MFA. In the end, flux inference from measured labeling enrichment is performed with a model encompassing biomass precursor withdrawal and all amino acid biosynthesis reactions. The focused model of section 8.2, though enabling exact predictions, does not allow to integrate a realistic measurement setup comprising label information from amino acid pools. Furthermore, the focused model assumes the CO_2 pool as an external parameter and only a comprehensive model is able to accurately capture the dependence of the CO_2 labeling and formation on the substrate isotopomer mixture (see Supplementary Information of Kappelmann et al. [90]).

For these reasons a numeric FIA was performed assuming partial information on a comprehensive model of the central metabolism of *C. glutamicum* as introduced in section 8.5. To globally determine the rDOF of the network topologies representing the wild-type, the Δ PEPCx and the Δ (PEPCx ME) strain the corresponding sensitivity matrices using the measurement system of Table S3 from the Supplementary Information of Kappelmann et al. [90] were analyzed for singularities on a sample of 1000 free flux sets.

The comprehensive carbon atom transition model of section 8.2 was implemented using 13CFLUX2 as modeling and simulation suite [188] and numerical simulations were performed on a substrate composition documented in the Supplementary Information of Kappelmann et al. [90]. Due to finite precision computations the rank of the Jacobian matrix of the comprehensive network $\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f}$ is based on a SVD which decomposes $\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f}$ according to:

$$\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad (4.3)$$

$\mathbf{\Sigma}$ contains singular values on the diagonal in descending order. A singular value in the order of machine precision represents a singular dimension of matrix $\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f}$. When $\text{rank}(\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f})$ is denoted by r and $r \leq \min(m, n)$, where m is the number of rows and n the number of columns of $\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f}$, the $(r + 1)$ -th to n -th columns of $\mathbf{V}$ span the nullspace of $\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f}$. Local identifiability holds, when the rDOF calculated from the rank of the Jacobian is zero at one given point $\mathbf{v}_{f,i}$ in the parameter space:

$$\text{rDOF} = \dim(\mathbf{v}_f) - \text{rank}\left(\frac{\partial \mathbf{y}(\mathbf{x}_{\text{inp}}, \mathbf{v}_{f,i})}{\partial \mathbf{v}_f}\right) = 0 \quad (4.4)$$

For global identifiability this relation has to hold over the entire domain of $\mathbf{v}_f$, i.e. for all possible admissible settings of free fluxes. Therefore, the singularity of the Jacobian was investigated for a set of uniform flux samples (see Supplementary Information of Kappelmann et al. [90]).

Figure 4.3 shows the distributions of the logarithm of the three smallest singular values of $\frac{\partial \mathbf{y}}{\partial \mathbf{v}_f}$ for the WT and the Δ PEPCx network topologies. It appears that the singular value spectrum separates into significant and non-significant singular values. It can be clearly seen that independently of the actual flux map, the sensitivity matrix of the WT background suffers from a rank deficiency of one (rDOF = 1), whereas the matrix describing the Δ PEPCx strain implies rDOF = 2.

Figure 4.4 A depicts the histogram arising for the Δ (PEPCx ME) network topology suggesting no rank deficiency (rDOF = 0). Again, the structural rank deficiency is independent of the actual flux map. This is not a surprising result since the analysis of structural contributions of fluxes to the equilibrium labeling fractions depends on the connectivity of the graph and not on the actual values of the fluxes. To show a more critical example, Figure 4.4 B depicts the arising histogram for the same network topology with the label information of SUC removed from the measurement model. Here no clear separation of singular values is obtained and the general limitation of the numerical FIA is exposed. Both limitations are circumvented by the integer algorithm presented above.

Although there are no rDOF for the Δ (PEPCx ME) network topology the uniqueness of the solution

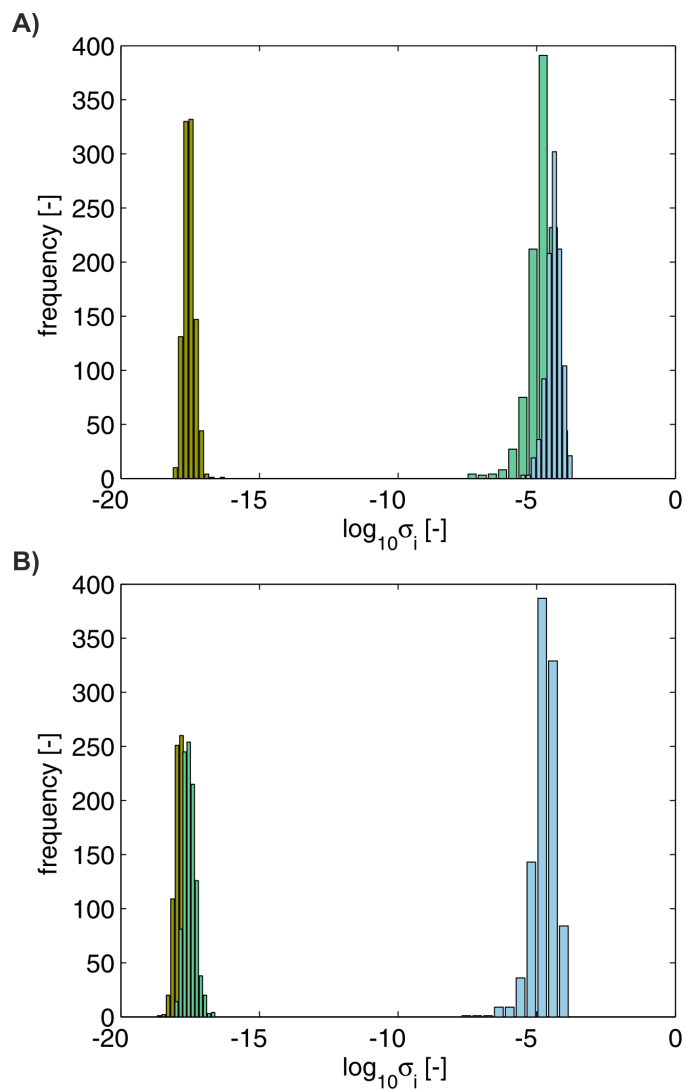


Figure 4.3: Global identifiability of central metabolic reactions in *C. glutamicum* under partial information. Histograms show the logarithm of the third last (blue), second last (green) and last (other) singular value of the sensitivity matrix evaluated on 1000 sets of free fluxes for the network topologies of the A) WT and B) ΔPEPCx strain.

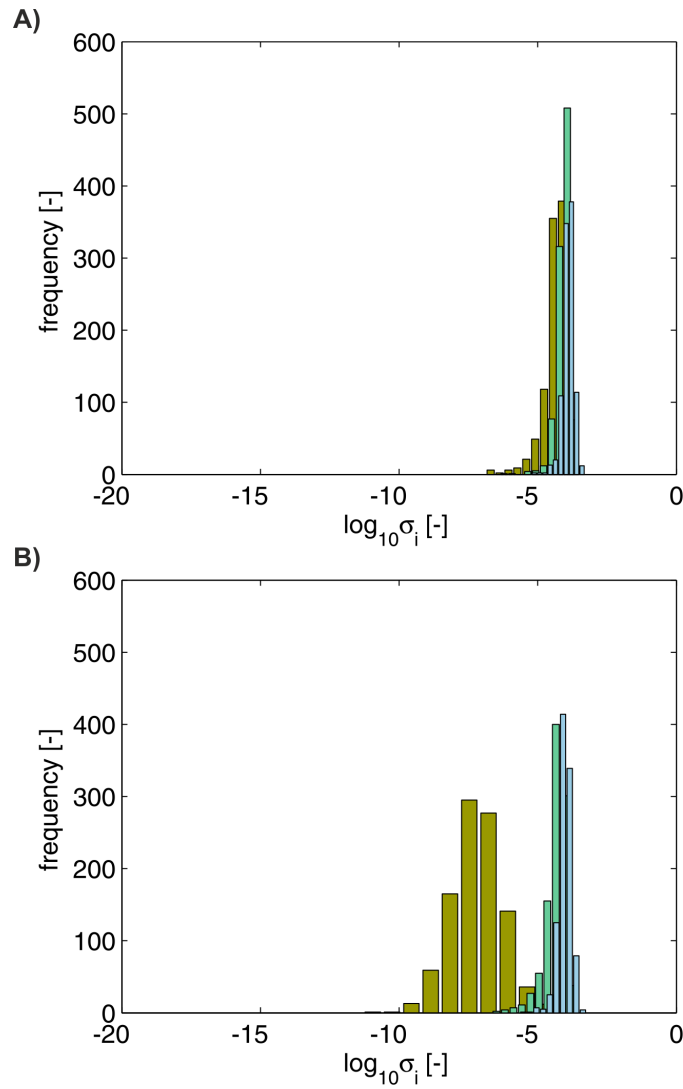


Figure 4.4: Global identifiability of central metabolic reactions in *C. glutamicum* under partial information. Histograms show the logarithm of the third last (blue), second last (green) and last (ocher) singular value of the sensitivity matrix evaluated on 1000 sets of free fluxes for the network topologies of the A) Δ (PEPCx ME) with the complete measurement system of Table S3 of the Appendix to Kappelmann et al. [90] and B) with the omission of SUC from the measurement system of Table S3 of Kappelmann et al. [90].

at every point of the histogram cannot be rigorously proven. Notwithstanding, for a sample of label distributions generated from a sample of free fluxes from the same domain as the one in Figure 4.4 a nonlinear regression from 100 randomly chosen starting points always converged to the true flux vector.

8.6 Conclusions

Summarizing, the predictions from the focused network model of the anaplerotic node, for which the structural non-identifiability was unambiguously proven, scale to a comprehensive model of *C. glutamicum* and show the validity of network simplifications represented by the focused model. Together, the applied approaches to identifiability analysis establish that the full set of anaplerotic reactions is non-identifiable and prove that the inflated confidence bounds in previous studies are an artefactual result arising from structurally non-identifiable fluxes.

9 Extracellular phenotyping of anaplerotic deletion mutants

9.1 Motivation

The identifiability analysis executed above yielded a set of flux-identifiable network structures which correspond to gene deletions in *C. glutamicum*. This finding motivated the construction of a set of deletion mutants, derived from one parental WT strain of *C. glutamicum*. All deletion mutants were constructed and kindly provided by Dr. Bastian Blombach, Institute for Biochemical Engineering, University of Stuttgart. The strains investigated in this thesis are listed in Table 4.3. The leftmost column therein lists the names which will be henceforth used to refer to the genotypes of the second column. The investigated set of strains comprises mutants for which a unique solution using ^{13}C -MFA can be found as well as mutants which are non-flux-identifiable strains. The latter were included to enable a comparison and plausibility control.

Table 4.3: Genotype and reference for the used strains in this work.

Strain	Genotype	Reference
<i>Corynebacterium glutamicum</i> WT	<i>C. glutamicum</i> ATCC13032	Blombach et al. [22]
<i>C. glutamicum</i> ΔPEPCx	Δppc	Blombach et al. [22]
<i>C. glutamicum</i> ΔPCx	Δpyc	Blombach et al. [22]
<i>C. glutamicum</i> ΔME	$\Delta malE$	Blombach et al. [22]
<i>C. glutamicum</i> ΔPEPCk	Δpck	Blombach et al. [22]
<i>C. glutamicum</i> $\Delta(\text{PEPCk ME})$	$\Delta(pck malE)$	This study
<i>C. glutamicum</i> $\Delta(\text{PEPCx ME})$	$\Delta(ppc malE)$	This study
<i>C. glutamicum</i> $\Delta(\text{PCx ODx})$	$\Delta(pyc odx)$	This study
<i>C. glutamicum</i> $\Delta(\text{PCx PEPCx})$	$\Delta(pyc ppc)$	This study

For all mutants cultivations in bioreactors were carried out to reliably estimate specific extracellu-

lar rates. In particular, the specific CO₂-production, glucose uptake and growth rate were estimated. This is an important step as the magnitude of the specific growth rate already gives a hint to the maximal anaplerotic activity. It can be expected that reduced anaplerotic activity directly reflects itself in reduced rates at which the biomass precursors GLU and ASP are formed (both being derived from the TCA cycle). The diminished biosynthesis of both should ultimately reflect itself in a lower specific growth rate. At the same time, the extracellular rates are a prerequisite for ¹³C-MFA, for which they need to be estimated reliably.

9.2 Adaptation of the $\Delta(\text{PCx PEPCx})$ strain to glucose as sole carbon source

All deletion mutants from Table 4.3 with the exception of the $\Delta(\text{PCx PEPCx})$ strain were able to grow on glucose without lag phase (data not shown). The $\Delta(\text{PCx PEPCx})$ mutant was able to grow without lag-phase on acetate, from which a cryo-culture was produced. Upon incubation on CGXII medium with glucose as sole carbon source, this strain initially showed no biomass growth. This observation is concordant with Peters-Wendisch et al. [143], who reported that a strain deficient in the corresponding genes showed no growth on glucose within a 48-hour interval. This prompted the authors to conclude that other potential anaplerotic enzymes in *C. glutamicum*, to wit ME, PEPCk, ICL and MS, are not capable of contributing to anaplerotic activity under glycolytic conditions. However, after prolonged incubation of cells from the $\Delta(\text{PCx PEPCx})$ cryo culture on CGXII medium with glucose, growth could be observed.

To precisely track the time-lag and growth behavior, a MTP plate was inoculated with the $\Delta(\text{PCx PEPCx})$ cryo-culture and incubated in the BioLector. The resulting time course of biomass growth for different inoculation densities is shown in Figure 4.5 A. Each subplot in this figure shows the growth of all wells of the same inoculation density, decreasing from top to bottom. Apparently, the overwhelming majority of cultures did adapt to glucose as sole carbon source after several days. This observation allows to formulate three different hypothesis: *I*) the growth on glucose of this mutant needs an adaptation in the form of expression changes of regulatory proteins and central carbon metabolism enzymes. Provided that this hypothesis is true, the metabolic rearrangement for growth on glucose accounts for the time needed till the onset of growth; *II*) Alternatively, a growth-conferring mutation already present in the population at the time of inoculation enables some cells to grow immediately or adapt biochemically. Since only cells bearing the corresponding mutation were able to grow, the real inoculation density (in terms of total number of cells able to proliferate) would be lower than the nominal inoculation density, but should nonetheless reflect the dilution of inoculum. *III*) Third, the time-lag represents the time needed for a certain mutation to occur that enables growth on glucose in the absence of PCx and PEPCx activity. In this case the realized adaptation through mutation can be expected to be different in each well based on the notion that a mutation represents a random variable with independent realizations in each well. Consequently, the time it takes for a culture to reach the same biomass concentration should be a random variable as well.

Figure 4.5 B shows the frequency distribution of the time each well needed to reach a Backscatter

value of 50. Each row corresponds to all those wells inoculated to the same OD. Under hypothesis *I*) the genotype of all cells in each well is growth-conferring from the beginning. After a biochemical adaptation, all cells should contribute to the Backscatter increase. Consequently, the time till a Backscatter of 50 is reached should closely reflect the inoculation densities assuming that the adaptation time is independent of the inoculation density ¹. As apparent from this figure, this cannot be confirmed since the lag-time does not reflect the decrease in inoculation density but rather appears to become more stochastic towards lower inoculation densities. This cannot be explained by cross-contamination of individual wells since the sterile control wells did not show an increase in biomass. On the other hand, if a mutation is needed for each culture (well) to adapt, each lag-time should be randomly distributed (hypothesis *III*). However, the wells corresponding to the two highest inoculation densities show a reproducible onset of growth, invalidating hypothesis *III*. Both observations can actually be reconciled by hypothesis *II*. If a growth-conferring mutation present at the start of the experiment makes up a small share of the inoculating cells, an increased relative variance in the number of cells able of growth in each well can be expected with increased dilution. This directly translates into increased variance in lag-time.

After growth had ceased, 200 μ l of well B4 ($OD_{init}=1$) was used to inoculate a subsequent shaking flask culture, from which a cryo-culture was produced. From this cryo-culture a sample was generated for whole-genome sequencing, which serves to substantiate the hypothesis formulated above. The sequencing and annotation of the polymorphisms and genetic variations was performed in the group of Tino Polen, IBG-1 of Forschungszentrum Jülich. The detected insertions and deletions are given in Table 4.4, whereas all identified single nucleotide polymorphisms (SNPs) are shown Table 4.5.

¹The time till a certain Backscatter value is reached is proportional to $\log(1/X_0)$, when the adaptation time is constant.

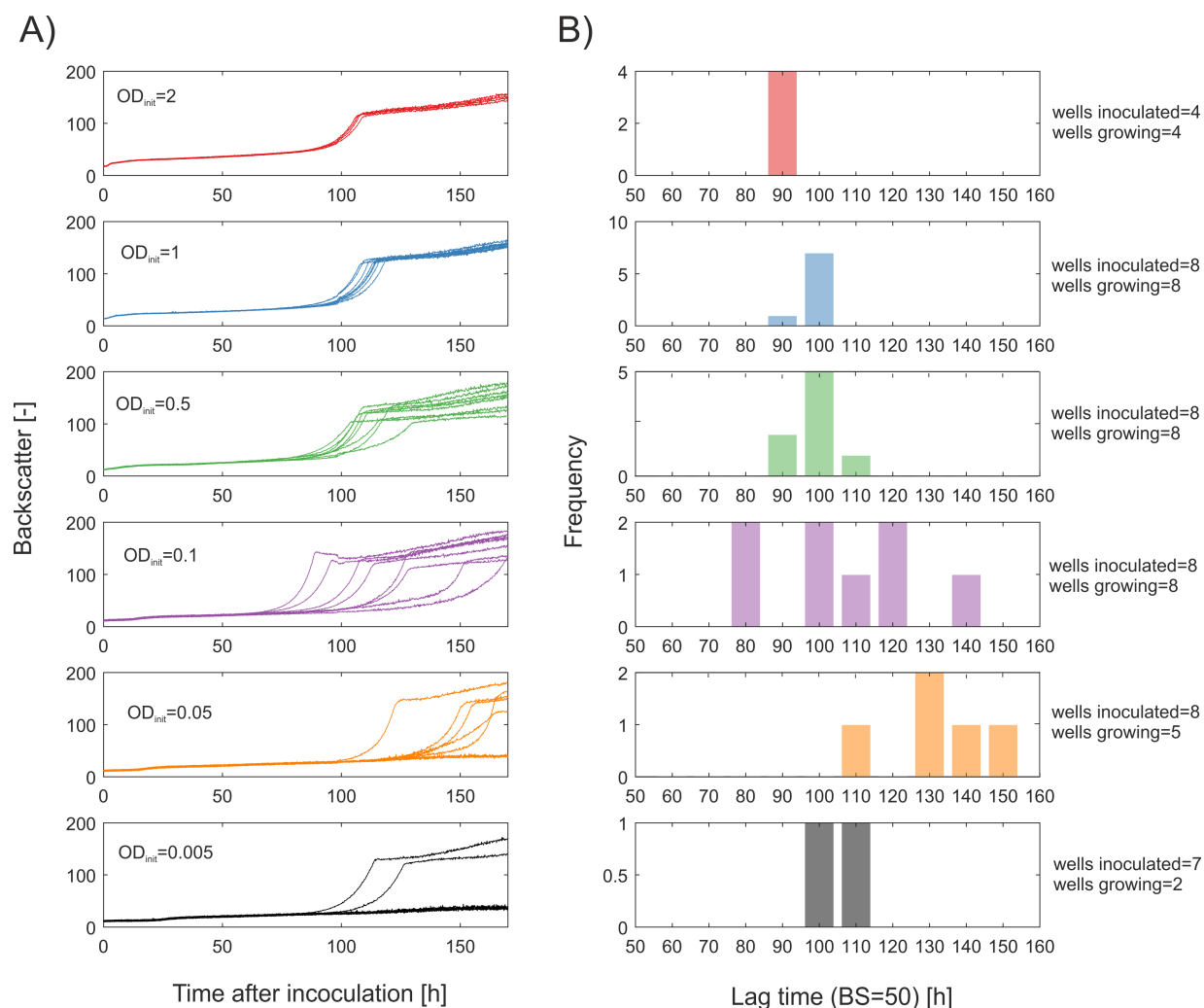


Figure 4.5: Phenotyping of the $\Delta(PCx\ PEPCx)$ mutant on 1 % glucose CGXII medium. A) Time course of biomass growth where each line represents one well of a certain inoculation density. The inoculation density is indicated on the left hand side of each plot. B) Histograms of the lag-times in each well of a certain inoculation density. Lag-time is defined as the time after inoculation needed to reach the Backscatter value of 50. In each plot the number of wells is given that were inoculated to the $OD_{init,600}$ indicated under A) alongside the number of wells that reached the Backscatter value of 50.

Table 4.4 reveals that the deleted activities of PCx and PEPCx are indeed absent at the population level. The corresponding deletions are identified with high confidence and show that the observed growth is not due to contamination. Table 4.5 reveals that several SNPs are present in this strain that are fixed in the population. The SNP in the intergenic region between cg3314 and cg3315 appears four nucleotides upstream from the translation start of MalR (*malR*, cg3315). This protein has been identified as repressor of the *malE* gene in the study of Krause et al. [104]. Since the mutation occurred outside the coding sequence of this regulator, only expression changes of the gene are able to affect metabolism. However, the specific protein abundance of MalR in this mutant appears unaffected in comparison to the WT (see section 11.5).

Further down the list, some SNPs affect the central carbon metabolism enzyme ICD. This enzyme

9. EXTRACELLULAR PHENOTYPING OF ANAPLEROTIC DELETION MUTANTS

Table 4.4: Identified insertions and deletions in the Δ (PCx PEPCx) strain adapted to glucose as sole carbon source in comparison to four different cultures of *C. glutamicum* WT (the four reference cultures of *C. glutamicum* originated from the research groups of Michael Bott, Jan Marienhagen, Julia Frunzke, and Tino Polen, all IBG-1 of Forschungszentrum Jülich). Shown are insertions and deletions that are not present in either of the comparison strains.

affected region	character	position	length	reads of position	rel. frequency
in cg2353 putative protein	deletion	2236684..2238317	1634	47	1
in cg2854, <i>tnp2c</i> transposase	deletion	2716280..2717915	1636	59	1
in cg0691, <i>groEL</i> 60 KDA chaperonin (protein CPN60 HSP60-)	deletion	610994..612446	1453	85	0.99
in cg0791, <i>pyc</i> Pyruvate carboxylase (EC:6.4.1.1)	deletion	707185..709613	2429	76	0.99
in cg1787, <i>ppc</i> Phosphoenolpyruvate carboxylase (EC:4.1.1.31)	deletion	1679298..1681219	1922	60	0.98
in intergenic region between cg1860 putative membrane protein and cg1861, <i>rel</i> ppGpp synthetase, ppGpp pyrophosphorylase (EC:2.7.6.5)	replacement	1754002..1754038	37	73	0.98
upstream of cg0756, <i>cstA</i> carbon starvation protein A	deletion	669585..669597	13	46	0.9
in cg2262, <i>ftsY</i> signal recognition particle GTPase	deletion	2144947..2144964	18	13	0.4
in cg0766, <i>icd</i> Isocitrate dehydrogenase (EC:1.1.1.42)	insertion	681078/681079	57	27	0.38
in cg0766, <i>icd</i> Isocitrate dehydrogenase (EC:1.1.1.42)	insertion	681136/681137	57	26	0.32
in cg0953, <i>mctC</i> monocarboxylic acid transporter	deletion	884479..884487	9	13	0.31

Table 4.5: Identified SNPs in the Δ (PCx PEPCx) strain adapted to glucose as sole carbon source in comparison to four different cultures of *C. glutamicum* WT (the four reference cultures of *C. glutamicum* originated from the research groups of Michael Bott, Jan Marienhagen, Julia Frunzke, and Tino Polen, all IBG-1 of Forschungszentrum Jülich). Shown are SNPs that are not present in either of the comparison strains or have a clearly distinguishable frequency. The column 'reads of position' refers to the number of sequencing events corresponding to the position of the polymorphism. The relative frequency of a SNP refers to the number of times this polymorphism was detected relative to the total number of reads of this position.

affected region	reads of position	rel. frequency of SNP
intergenic region		
LEFT: cg3314, conserved hypothetical protein RIGHT: cg3315, transcriptional regulator malR repressor of the malic enzyme gene <i>malE</i>	181	1
exchange L328S in cg1245(+) putative membrane protein	102	1
T306T in cg1574(+), <i>pheS</i> phenylalanyl-tRNA synthetase a subunit EC:6.1.1.20	116	1
intergenic region		
LEFT: cg2136 <i>gluA</i> , glutamate uptake system RIGHT: cg2137 <i>gluB</i> , glutamate uptake system, secreted glutamate-binding protein	125	1
A22A in cg3237(+), <i>sodA</i> (sod) manganese superoxide dismutase EC:1.15.1.1	125	1
exchange Y158H in cg1676(-), <i>lip3</i> putative lipase EC:3.1.1.3	106	0.99
exchange E324A in cg1451(+) <i>serA</i> , phosphoglycerate dehydrogenase EC:1.1.1.95	75	0.35
exchange G577S in cg0766(-), <i>icd</i> , isocitrate dehydrogenase EC:1.1.1.42	148	0.33
exchange K485E in cgr01(+) 16S ribosomal RNA	6	0.17
exchange P583S in cg0766(-) <i>icd</i> , isocitrate dehydrogenase (EC:1.1.1.42)	149	0.14
Stop L508* in cg2267(-) putative membrane protein	89	0.11
exchange V117E in cg0237(-) putative short chain dehydrogenase or oxidoreductase	76	0.11
exchange K922R in cgr17(-) 23S ribosomal RNA	10	0.1

catalyzes the first reaction of the oxidative decarboxylation branch of the TCA cycle, which is one of two metabolic routes emanating from the ICIT branch point. An alternative route for ICIT from this branch point consists in the conversion by ICL to GLX and SUC, which constitutes the first reaction of the GLX shunt. Overall, the reactions of the GLX shunt synthesize one equivalent of a C4-body per two equivalents of ACCOA and thus represent one anaplerotic reaction sequence. In light of this, it appears likely that the identified genetic alterations in the *icd* gene may negatively affect the function or affinity of ICD towards ICIT, which would reduce the carbon flux through the oxidative decarboxylation. The reverse hypothesis of increased ICD activity appears unlikely as it would increase the decarboxylation of the carbon skeleton of ICIT and reduce anaplerotic activity. However, this mutation was only identified in a fraction of all reads of the corresponding nucleotide position. If the total number of reads is representative of the total cell population, only a third of all cells exhibit this mutation. Since a single sample from one time point was analyzed, it is not clear whether this mutation was at the point of enriching itself further in the population or whether it represents a stable optimum in an adapted culture. Further time series experiments are needed to validate this hypothesis. Furthermore, two insertions were detected in this enzyme (Table 4.4) at approximately the same share as the SNP affecting this enzyme.

The annotations of the remaining ORFs do not allow to draw a clear picture of the effects of the resulting amino acid exchanges for the affected ORFs are not part of the central carbon metabolism or transcriptional regulatory network. The deletion detected in the *mctC* gene (cg0953) and its consequences for the exometabolome will be discussed later in this chapter.

9.3 Specific extracellular rates of anaplerotic deletion mutants

The entire set of anaplerotic deletion mutants was cultivated under controlled bioreactor conditions on defined CGXII medium with 1% glucose in biological duplicate or triplicate. For two representative cultivations of the Δ PEPCk and Δ (PCx PEPCx) strains the time course of biomass, substrate and CO₂ signals is given in Figure 4.6A and B, respectively. To estimate specific extracellular rates from these data, a model-based approach was chosen (see Material and Methods chapter). A model for data integration allows to reconcile and harvest all data that ought to reflect the specific growth rate for rate estimation. In particular, the biomass and substrate signals were fitted using the set of equations (3.6) - (3.7), yielding the specific uptake and growth rate under the assumption of constant specific rates during the exponential phase. The time course of the fitted signals is shown in Figure 4.6 as dashed lines. The substrate uptake rate of 4.82 [CI: 4.61-5.07] mmol (g CDW)⁻¹ for *C. glutamicum* WT determined by this approach agrees well with the one determined in the study of Buchholz et al. [24], who reported 5.056 mmol (g CDW)⁻¹ h⁻¹ as uptake rate on glucose for the WT. The estimate of the growth rate of 0.45 h⁻¹ also agrees well with literature. Besides the specific rates of the WT, the determined extracellular rates for the entire set of anaplerotic deletion mutants are shown in Figure 4.7. Based on the bootstrapped confidence intervals discernible growth phenotypes can be identified among the set of deletion mutants.

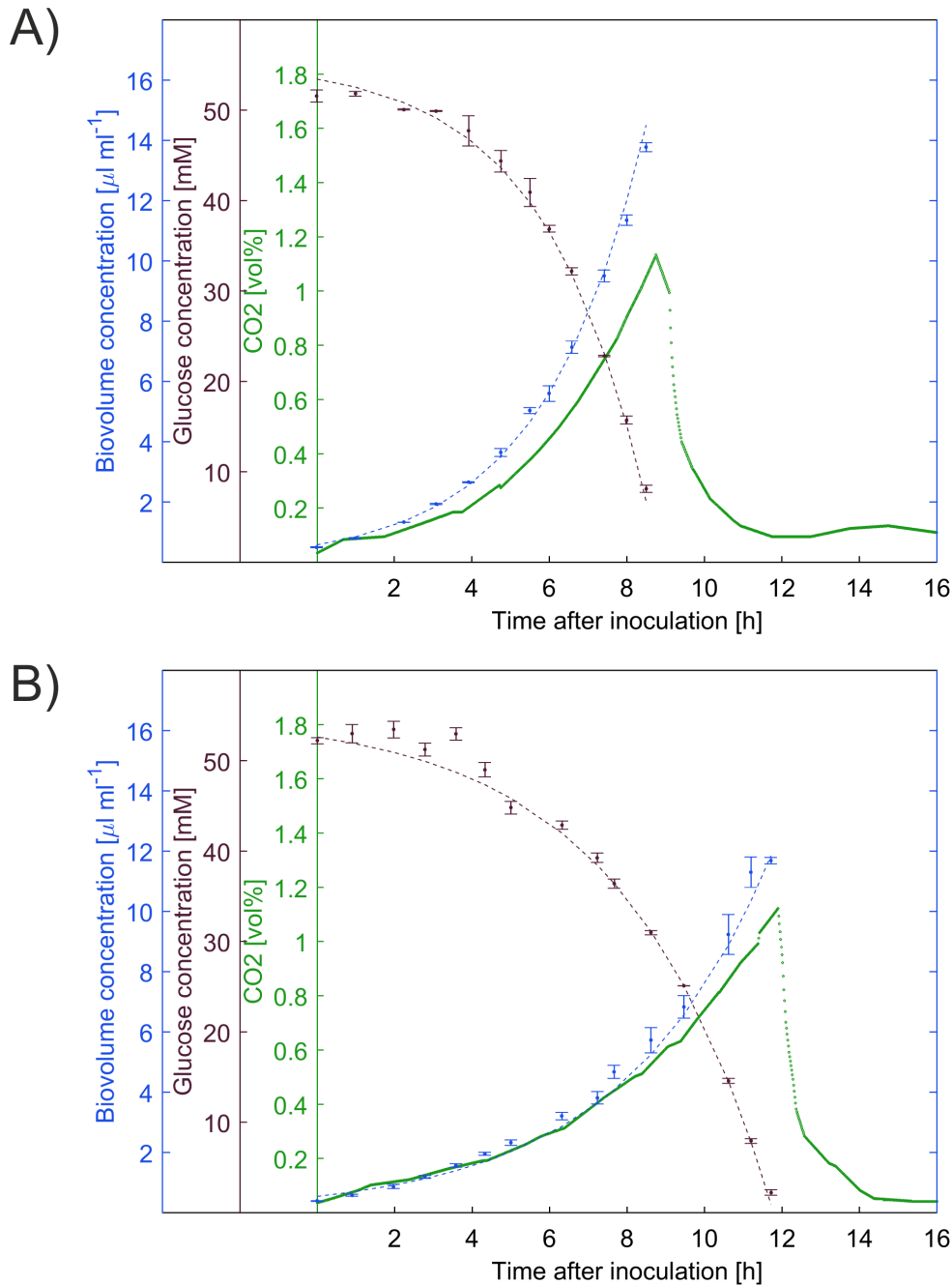


Figure 4.6: Time course of BV and substrate concentration as well as CO₂ exhaust gas volume fraction from a cultivation of the Δ PEPCk mutant (A) and the Δ (PCx PEPCx) mutant (B) on defined CGXII medium with 1% glucose. BV and glucose concentrations are indicated together with their standard deviations from replicate measurements. Dashed lines show the computed time course of the latter two signals for the best fit of equations (3.6) and (3.7) to the shown experimental time course. Green filled circles denote the time-dependent volume fraction of CO₂ in the exhaust gas stream.

First, the PEPCk mutant has a lower growth rate although PEPCk catalyzes a gluconeogenic reaction, not needed under glycolytic conditions to supply biomass synthesis precursors. However, its growth defect clearly indicates its activity under glycolytic conditions. This result was also obtained in

the study of Petersen et al. [145], who inferred PEPCK activity in *C. glutamicum* WT from ^{13}C -labeling data. Since the carbon flow catalyzed by PEPCK is of opposite direction as the overall carbon flow under glycolytic conditions, its function must be of indirect nature. Curiously, the double deletion mutant $\Delta(\text{PEPCK ME})$ shows a restored growth phenotype, with a growth rate equivalent to the WT. At this point, no conclusion with respect to this effect can be made.

Deletion mutants comprising a deletion in PEPCK do not show a significantly altered growth phenotype so long as PCx is still active. The stoichiometry of the reaction catalyzed by PEPCK is identical to the reaction sequence of PK and PCx which apparently fully compensates for the missing carboxylation activity since no growth defect is apparent in the $\Delta(\text{PEPCK})$ and $\Delta(\text{PEPCK ME})$ mutants. On the other hand, mutant strains ΔPCx and $\Delta(\text{PCx ODx})$ do exhibit a growth phenotype. This finding underscores the role of PCx as most important anaplerotic reaction and already suggests that its catalyzed flux is most likely greater than that of PEPCK in the WT.

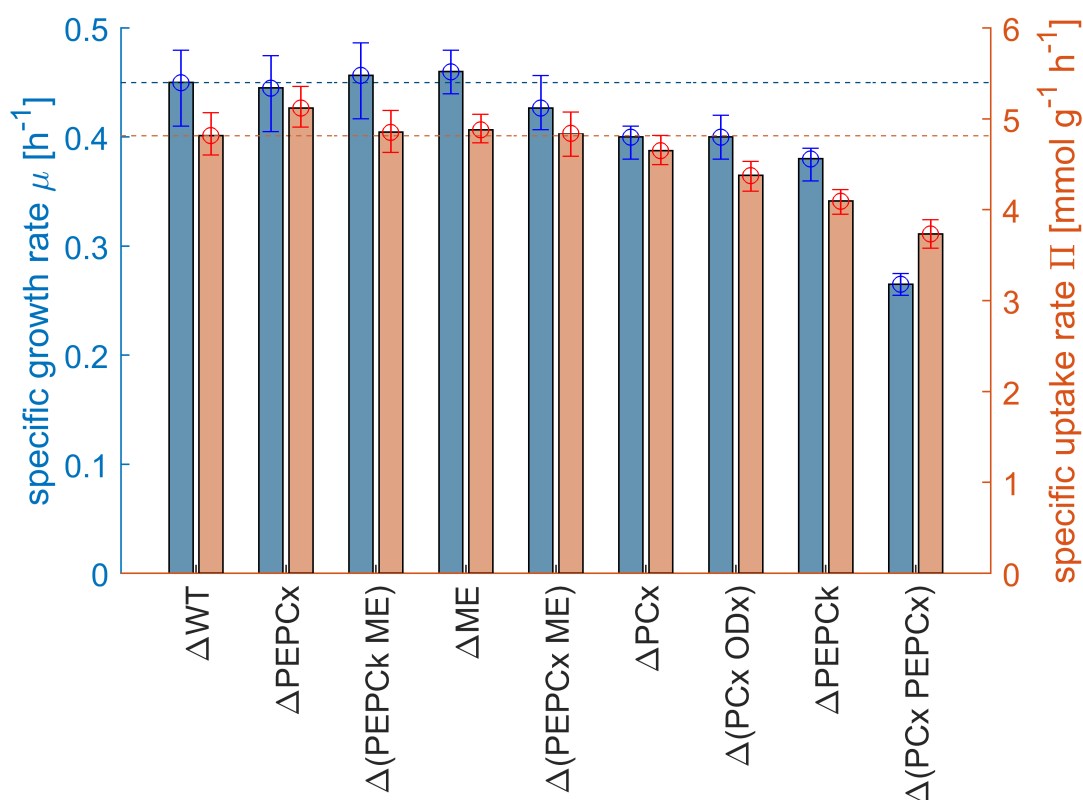


Figure 4.7: Estimated specific growth and uptake rate for anaplerotic deletion mutants. To integrate estimated rates from several replicate cultivations of the same strain, Figure 4.7 shows the mean values of parameter estimates from replicate cultivations. The proposed rate estimation scheme determines confidence bounds on each estimate from one cultivation by bootstrapping according to section 7.10. The error estimates in Figure 4.7 are the minimum of the lower bound and the maximum of the upper bound of each parameter over all replicate cultivations of the mutant in question.

The $\Delta(\text{PCx PEPCKx})$ mutant, missing both carboxylation activities shows a greatly reduced mean

growth rate of 0.27 h^{-1} . In the absence of both PCx and PEPCx, three different anaplerotic activities can possibly substitute the anaplerotic activity: First, ME may catalyze the carboxylation of PYR to MAL in an NADPH-dependent manner. This appears plausible since ME was found to catalyze this reaction sequence in *in vitro* assays [33, 63]. Moreover, it could be experimentally shown that a NADPH-dependent ME from *E. coli* can act as sole anaplerotic enzyme in *Saccharomyces cerevisiae* sustaining a growth rate of 0.06 h^{-1} [207]. Second, PEPCk may act in reverse direction from PEP to OAA. This hypothesis is rather unlikely as this reaction directionality would couple anaplerotic carboxylation to the substrate level phosphorylation, generating one GTP molecule. Since the reaction from OAA to PEP is coupled to the hydrolysis of one GTP molecule, it can be expected to be favored. A third alternative represents the GLX shunt, which exclusively fulfills the anaplerotic function in *C. glutamicum* under growth on acetate [189]. The sequencing results from above suggest an altered ratio of ICD and ICL activity at the ICIT branch point and therefore favor the hypothesis of an active GLX shunt contributing to anaplerosis in this mutant.

The $\Delta(\text{PCx PEPCx})$ mutant also shows an altered ratio of specific glucose uptake and growth rate in comparison to other mutants and the WT. The observation that less biomass is formed per unit uptake rate in comparison to the other strains raises the question as to where the excess carbon atoms end up. To answer this question, a carbon balance was computed. Contrary to the conventional carbon balancing approach, consisting in the quantification of the total carbon recovery in biomass and exhaust gas by the time all substrate has been consumed [24], the carbon balance in this work was computed according to equation (4.5). This equation balances the specific rates of carbon uptake and carbon flow into sinks at any given time. This point of view is relevant for ^{13}C -MFA, since the instantaneous specific rates are used in the flux estimation and therefore need to account for all carbon fluxes.

$$\Theta = \frac{\frac{(\mu * \omega_{\text{Carbon}} * 1000)}{M_{\text{Carbon}}} + \Pi_{\text{CO}_2, \text{spec}}}{\Pi_{\text{Glc, spe}} * 6} \quad (4.5)$$

The numerator of equation (4.5) sums the specific CO_2 -formation rates and the carbon equivalent of the specific growth rate. Both quantities are the only considered carbon sinks in the network of section 8.2. Therefore, a closed balance of the specific carbon uptake rate and the rate at which carbon flows into sinks is an important validation of the proposed model in section 8.2.

To compute the carbon equivalent of the specific growth rate, the mass fraction of carbon in the CDW, ω_{Carbon} , needs to be known. This quantity has been reported several times in the literature: Marx et al. [117] reported 0.408 for *C. glutamicum* MH20-22B, a lysine producer strain, determined in freeze-dried biomass with a CHNS elemental analyzer. More recently, Buchholz et al. [24] reported a value of 0.514 for ω_{Carbon} in *C. glutamicum* ATCC 130322. This value was determined by separately quantifying the total carbon in liquid bioreactor samples (supernatant and biomass) and the total inorganic carbon (total dissolved carbon) in the supernatant. An orthogonal access to ω_{Carbon} may come from concentration measurements of individual biomass components. This information is in-

incorporated in the biomass formation equation of the comprehensive model of *C. glutamicum* central metabolism from section 8.2. In this model the concentrations of all biomass components in mmol (g CDW)⁻¹ translates into a value for ω_{Carbon} of 0.466. In the course of this work, the carbon content in CDW was experimentally quantified by the Central Institute for Analytics (ZEA-3) of Forschungszentrum Jülich to be around 0.4-0.42. The uncertainty of the parameter ω_{Carbon} notwithstanding, its value was assumed to be 0.45 for all following calculations. In face of the apparent uncertainty, its standard error was assumed to be 0.05. The arising interval covers all available literature information on this quantity. The uncertainty in derived quantities thereof can be computed by Gaussian error propagation.

The numerator and denominator of equation (4.5) are shown separately in Figure 4.8 for each strain. From this figure it becomes apparent that the $\Delta(\text{PCx PEPCx})$ strain grows with a higher relative CO₂-formation rate with respect to the uptake rate. At the same time, the sum of the specific rates at which carbon flows into sinks amounts to the same relative value relative to the carbon uptake rate as in other strains. Therefore, it seems plausible that a higher decarboxylation activity explains the lower relative growth rate in the $\Delta(\text{PCx PEPCx})$ strain.

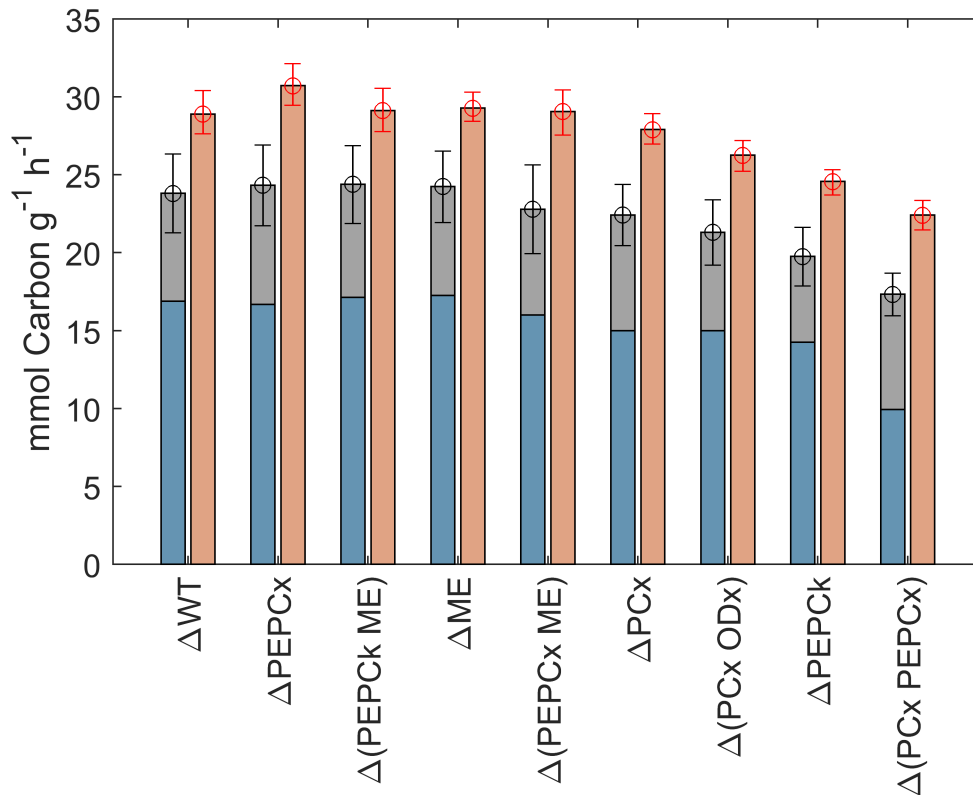
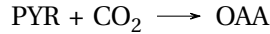
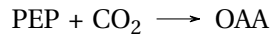


Figure 4.8: Carbon equivalents of the specific growth rate (blue), CO₂-formation rate (gray) and the substrate uptake rate (orange). The error bar on the stacked bars is the error on the sum of both rates. This error estimate was computed according to Gaussian error propagation using the equations of Appendix 4 and mainly derives from the uncertainty in ω_{Carbon} .

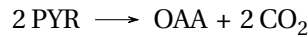
A higher relative CO₂-formation rate already substantiates the hypothesis that an altered ratio of ICL and ICD activity involves the GLX shunt as anaplerotic reaction sequence in the Δ(PCx PEPCx) strain. Anaplerotic activity through ME or a reversible PEPCK enzyme would exhibit the same CO₂ balance as the carboxylation catalyzed by PCx and PEPCx per C4-body formed. Both reactions fix one equivalent CO₂ per OAA equivalent formed. For PCx and anaplerotic ME, the reaction is



For a reversible PEPCK the same carbon balance as for PEPCx holds true:



For the GLX shunt, on the other hand, the following balance can be set up:



Thus, exclusive anaplerotic activity through the GLX shunt would release two equivalents CO₂, instead of fixing one per C4-body formed. Thus, based on the extracellular phenotyping alone, it seems plausible to assume that the GLX shunt at least partly contributes to the anaplerotic activity in the Δ(PCx PEPCx) strain. Surely, altered reaction rates of other CO₂-dependent reactions may also affect the observed specific CO₂-production rate. In particular, an increase in G6PD activity may also contribute to an increased CO₂-release. However, it is unlikely that a change in the rate of this reaction in the Δ(PCx PEPCx) strain accounts for the altered CO₂-production rate, given the magnitude of the difference with respect to the WT.

It has been assumed that the GLX shunt genes in *C. glutamicum*, *aceA* and *aceB*, are repressed under glycolytic conditions and de-repressed when acetate is metabolized [12]. When the latter is the sole carbon source, the GLX-shunt represents the only reaction sequence for anaplerosis in *C. glutamicum*. Conspicuously, the growth rate of the Δ(PCx PEPCx) strain lies in the range of the reported growth rate in the study of Wendisch et al. [189] for growth of *C. glutamicum* WT on acetate as sole carbon source. This gives an indication that the GLX-shunt may very well be the only active anaplerotic reaction in this strain. Indeed, as will be shown later all intracellular measurements suggest the same.

The estimated specific extracellular rates already allow to derive constraints as to the prevailing flux distribution in some strains. Since the ΔPEPCK mutant exhibits a clear growth phenotype, the corresponding reaction must be active in the WT. Otherwise, its removal would be inconsequential. The same holds true for ME in the ΔPEPCK mutant, since its removal restores the normal growth phenotype of the Δ(PEPCK ME) mutant. Lastly, PEPCx must catalyze a biologically significant flux when PCx is deleted since the Δ(PCx PEPCx) strain shows a severe growth phenotype in comparison to the ΔPCx mutant. These findings provide a benchmark to which the flux estimation in a later section must live up.

9.4 Bicarbonate pool as carbon sink

One result holds true irrespective of the genetic background: The recovery of carbon at any given time in the reactions that act as carbon sinks amounts to 80 % of the carbon equivalent of the uptake rate (see Figure 4.8). This ratio is denoted Θ and computed according to equation (4.5). It will henceforth be called instantaneous carbon balance. The value for Θ in each cultivation is listed in Table 2 of the Appendix to this thesis.

The non-closed instantaneous carbon balance may hint to extensive by-product formation or indicate systematically biased extracellular rates. The former is unlikely since *C. glutamicum* WT is known to produce only minor by-products under aerobic conditions (the DO was maintained at 30 %). Notwithstanding, the genetic alterations may induce a more extensive overflow metabolism in some deletion mutants. However, as will be demonstrated in section 10.2 of this chapter all organic acids, sugar phosphates and amino acids in the culture supernatant measured by targeted LC-MS/MS account for a total of 274 - 786 $\mu\text{mol carbon (g CDW)}^{-1}$, depending on the strain. Although the exometabolome indeed differs between the strains, it can be neglected as carbon sink since the gap in the balance of specific rates amounts to several $\text{mmol carbon (g CDW} \cdot \text{h)}^{-1}$ (see Figure 4.8).

Remarkably, the value of 80 % for the instantaneous carbon balance matches the determined carbon balance closure in the study of Buchholz et al. [24]. This study convincingly showed that the gap in carbon balance can be traced back to the fact that part of the CO_2 -production at any given time dissolves as HCO_3^- in the culture broth. This share is not recovered as gaseous CO_2 at the detector, leading to an underestimation of the CO_2 -production rate. However, this study quantified this effect in 100 l bioreactors with a working volume of 25 l, which is too great a difference in culture volume to the present thesis to directly translate this finding. Higher working volumes entail a higher average hydrostatic pressure in the liquid phase and consequently a higher CO_2 solubility. Therefore, the bicarbonate formation and its consequences for the estimation of the specific CO_2 -formation rate for ^{13}C -MFA will be studied in more detail in the bioreactor setup used in this thesis.

The total experimental carbon recovery in form of CO_2 in the exhaust gas is quantified by $\hat{\Pi}_{\text{CO}_2, \text{tot}}(t)$ according to equation (3.8) from the Materials and Methods chapter, which arises from balancing the gas-phase of the bioreactor. Following the convention from above hat notation is used to differentiate the true quantity $\Pi_{\text{CO}_2, \text{tot}}(t)$ from the experimental measurement thereof. Furthermore, to keep the following mass balances as simple as possible, the amount of CO_2 is balanced in mol instead of mmol as above such that $\hat{\Pi}_{\text{CO}_2, \text{tot}}(t)$ is quantified as total CO_2 -formation in mol h^{-1} . Dividing the total CO_2 -formation rate by the liquid bioreactor volume from which it evolves yields the volumetric CO_2 -production rate $\hat{\Pi}_{\text{CO}_2, \text{vol}}(t)$ in mol (l h)^{-1} . Conceptually, the true value of this quantity can be related to the specific CO_2 -formation rate $\Pi_{\text{CO}_2, \text{spec}}$ during batch cultivations in the following way:

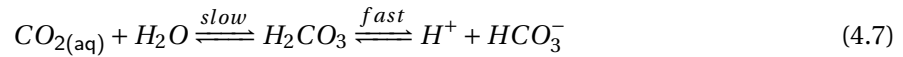
$$\Pi_{\text{CO}_2, \text{vol}}(t) = \Pi_{\text{CO}_2, \text{spec}} X_0 \exp(\mu t) \quad (4.6)$$

The term $X_0 \exp(\mu t)$ denotes the time-dependent biomass concentration in the culture volume.

The quantity $\Pi_{\text{CO}_2, \text{spec}}$ is given in $\text{mol (g CDW h)}^{-1}$ and can be assumed to be constant under the assumption of a quasi-steady state during the exponential phase of a batch cultivation.

For the calculated $\bar{\Pi}_{\text{CO}_2, \text{vol}}(t)$ to be equal to the true underlying $\Pi_{\text{CO}_2, \text{vol}}(t)$ at every time point during the cultivation, all CO_2 produced by the organism must be recovered instantaneously at the exhaust gas detector for it to enter equation (3.8) in the form of $\Phi_{\text{CO}_2}^o(t)$. However, when CO_2 diffuses through the cell wall, it will dissolve as $\text{CO}_{2(\text{aq})}$ in the medium and may take two possible pathways: Either it enters the gas phase via liquid-to-gas mass transfer leading to a signal at the exhaust gas detector or it forms bicarbonate through hydration.

In an aqueous system $\text{CO}_{2(\text{aq})}$ can be hydrated by two possible reactions, both of which are second-order reactions with either water or hydroxide ions [59]. The reaction with water proceeds as follows:



The dissociation step of carbonic acid H_2CO_3 to H^+ and HCO_3^- is assumed to proceed fast in comparison to the hydration step in equation (4.7). This assumption is justified since the dissociation step of carbonic acid is several orders of magnitude faster than the hydration [111]. For this reason just one rate constant needs to be considered, i.e. the rate constant of the rate-limiting hydration step. Under alkaline conditions $\text{CO}_{2(\text{aq})}$ can react with hydroxide ions to form bicarbonate directly:



From these reaction schemes it becomes apparent that the back-reaction from equation (4.7) and the forward-reaction from equation (4.8) are pH-dependent second-order reactions. However, since the pH is normally controlled at a constant value during the cultivation, the concentration of OH^- or H^+ ions are constant and can be absorbed into the rate constants of the respective reactions. The same holds true for the water concentration which is in great excess of all other components. Thus, a reversible pseudo-first order reaction for both reactions (4.7) and (4.8) with individual pH-dependent rate constants can be considered. This approach is already described in Gibbons and Edsall [59] and Johnson [84]. Both reactions can be lumped by adding the expressions for their rate constants. This leaves just a single reversible reaction with one pH-dependent forward and one pH-dependent backward rate constant. This, in turn, allows to express the total volumetric rate of bicarbonate formation in $\text{mol h}^{-1} \text{ l}^{-1}$ as:

$$r_{\text{HCO}_3^-} = k^+ [\text{CO}_{2(\text{aq})}] \quad (4.9)$$

The total volumetric $\text{CO}_{2(\text{aq})}$ formation in $\text{mol h}^{-1} \text{ l}^{-1}$ from HCO_3^- can be modeled as:

$$r_{\text{CO}_{2(\text{aq})}} = k^- [\text{HCO}_3^-] \quad (4.10)$$

The expressions for the lumped rate constants k^+ and k^- are pH-dependent and the values for pH

= 7 are listed in Table 3. The expressions for k^+ and k^- are taken from Johnson [84] who determined the rate constants in sea water. These are in agreement with experimental determination in water in the studies of Ho and Sturtevant [75] as well as Gibbons and Edsall [59]. The values of Table 3 are piece-wise linear interpolated to a temperature of 30°C using the data given in Johnson [84]. The pH-dependence of the forward and backward rate constants is depicted in Figure 4.9. It can be seen that the rate of reaction from HCO_3^- to $CO_{2(aq)}$ decreases exponentially with higher pH values. Thus, a system maintained at pH 8 will maintain a much higher bicarbonate concentration in comparison to a system a pH 7 when chemical equilibrium is reached between HCO_3^- and CO_2 . Although the shown rate constants are valid in sea water, they will at least approximate the values these constants take in CGXII medium containing a given biomass concentration of *C. glutamicum* and will surely serve to gain a qualitative understanding of the CO_2 -dynamics explaining the observed gap in instantaneous carbon balance.

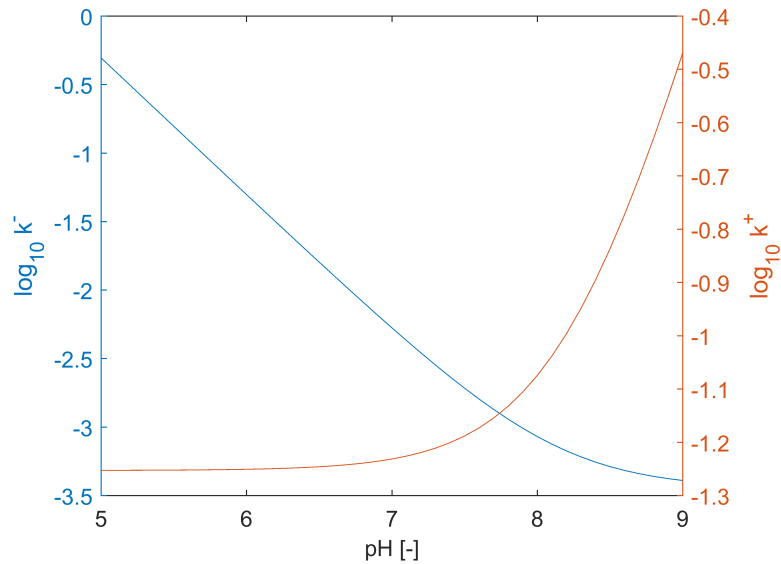


Figure 4.9: pH-dependence of the forward and backward rate constants of the hydration of CO_2 in sea water at 30°C.

The other pathway for the $CO_{2(aq)}$ species consists in mass transfer to the gas phase. This mass transfer is proportional to the difference between the current $CO_{2(aq)}$ concentration and the equilibrium concentration at the current gas phase. The equilibrium concentration $CO_{2(aq)}^*$ can be estimated by the following formula:

$$CO_{2(aq)}^*(t) = \frac{pCO_2(t)}{K_{H,CO_2}} \quad (4.11)$$

where $pCO_2(t)$ denotes the current gas phase partial pressure of CO_2 in Pa and K_{H,CO_2} the Henry's law volatility constant in $Pa \cdot l \cdot mol^{-1}$. The latter constant was obtained from the compilation of Sander [157]. Given these definitions, the total volumetric mass transfer $r_{gas \rightarrow liquid}(t)$ in $mol \cdot (l \cdot h)^{-1}$ can be

expressed in dependence of the overall mass transfer coefficient $k_L a$. $k_L a$ is the proportionality constant known from oxygen transport problems. However, this $k_L a$ constant is specific for CO_2 :

$$r_{\text{gas} \rightarrow \text{liquid}}(t) = k_L a \left(\frac{p\text{CO}_2(t)}{K_{\text{H},\text{CO}_2}} - [\text{CO}_{2(\text{aq})}](t) \right) \quad (4.12)$$

In summary, two reactions, i.e. mass transfer and hydration, compete for $\text{CO}_{2(\text{aq})}$ which is produced by biological decarboxylation activity. Assuming that the volume in which these reactions take place is constant, the following mass balance can be derived for the $\text{CO}_{2(\text{aq})}$ pool:

$$\frac{d[\text{CO}_{2(\text{aq})}](t)}{dt} = \Pi_{\text{CO}_2, \text{spec}} X_0 \exp(\mu t) + k_L a \left(\frac{p\text{CO}_2(t)}{K_{\text{H},\text{CO}_2}} - [\text{CO}_{2(\text{aq})}](t) \right) - k^+ [\text{CO}_{2(\text{aq})}] + k^- [\text{HCO}_3^-] \quad (4.13)$$

The term $\Pi_{\text{CO}_2, \text{spec}} X_0 \exp(\mu t)$ is the volumetric CO_2 -formation rate a time t . It may be noted that the mass transfer term evaluates to a negative value, when $[\text{CO}_{2(\text{aq})}]$ is greater than the equilibrium concentration at the current gas phase $p\text{CO}_2(t)/K_{\text{H},\text{CO}_2}$. Under these conditions an outflow from $[\text{CO}_{2(\text{aq})}]$ occurs through the negative sign of the mass transfer term. Each summand in equation (4.13) has the unit $\text{mol} (\text{l h})^{-1}$. Although k^- and k^+ are reported in s^{-1} in Johnson [84], they are used in h^{-1} in equation (4.13), since the inflow into $\text{CO}_{2(\text{aq})}$ through $\Pi_{\text{CO}_2, \text{spec}} X_0 \exp(\mu t)$ changes on the time scale of hours.

For the bicarbonate pool the mass balance is given by:

$$\frac{d[\text{HCO}_3^-](t)}{dt} = k^+ [\text{CO}_{2(\text{aq})}](t) - k^- [\text{HCO}_3^-](t) \quad (4.14)$$

We furthermore balance the amount of CO_2 in the gas phase by assuming one outflow through the exhaust gas stream and two inflows into the gaseous CO_2 pool:

$$\frac{V_{\text{air}}}{RT} \frac{dp\text{CO}_2(t)}{dt} = -k_L a \left(\frac{p\text{CO}_2(t)}{K_{\text{H},\text{CO}_2}} - [\text{CO}_{2(\text{aq})}](t) \right) V_{\text{liq}} - \frac{F_{\text{air}} p\text{CO}_2(t)}{RT} + \frac{P_{\text{ges}} \Phi_{\text{CO}_2}^\alpha / 100 F_{\text{air}}}{RT} \quad (4.15)$$

The mass transfer term in equation (4.15) is scaled by the liquid volume to obtain the total amount of mass transfer in mol h^{-1} . Thus, in equation (4.15) all summands have the unit mol h^{-1} . It may be furthermore noted, that the mass transfer term evaluates to a positive value when $[\text{CO}_{2(\text{aq})}](t)$ is greater than $p\text{CO}_2(t)/K_{\text{H},\text{CO}_2}$.

For the numerical integration of the resulting ODE system of equations (4.13), (4.14) and (4.15) the initial conditions are chosen to represent an aqueous phase in equilibrium with a gas phase of atmospheric CO_2 content and a bicarbonate concentration in equilibrium with the dissolved carbon dioxide. The parameters needed are the Henry's law volatility constant K_{H,CO_2} , $k_L a$ as overall mass transfer coefficient as well as k^+ and k^- as forward and backward rate constants for the reaction from dissolved CO_2 to HCO_3^- . For the Henry constant of CO_2 in water a temperature of 30°C was assumed. The values of all parameters can be found in Table 3 of the appendix To simulate the exponentially

increasing CO_2 -production with subsequent stop when substrate-limitation is reached the following formula for $\Pi_{\text{CO}_2, \text{vol}}$ was used:

$$\Pi_{\text{CO}_2, \text{vol}} = \Pi_{\text{CO}_2, \text{spec}} \cdot X_0 \cdot \exp(\mu \cdot t) \cdot \left[1 - S\left(\frac{t-8}{0.4}\right) \right] \quad (4.16)$$

where $S(t)$ denotes the sigmoid function. This function simulates "s"-shaped curve with a value range of [0,1]. Through the difference in equation (4.16) it describes an "s"-shaped decline such that the expression in equation (4.16) simulates an exponential increase of $\Pi_{\text{CO}_2, \text{vol}}$ up to 8 hours from $t = 0$ and then vanishes smoothly to zero. The introduction of sigmoid function is necessary to enable numeric integration. For $\Pi_{\text{CO}_2, \text{spec}}$ a value of $0.01 \text{ mol (g CDW h)}^{-1}$ was assumed as influx into the $\text{CO}_{2(\text{aq})}$ -pool.

Figure 4.10 shows the time course of bicarbonate and dissolved CO_2 pools for different parameter values. Figure 4.10 A depicts the time course of both pools with rate constants k^+ and k^- at pH=7. A clear accumulation of bicarbonate is visible resulting in an underestimated CO_2 -production rate which is depicted alongside to the right. Setting the forward rate constant k^+ to zero and thus impeding any bicarbonate formation results in the time course depicted in Figure 4.10 B. Expectedly, the true biological CO_2 -formation rate matches the calculated one almost exactly in this case. The match is not exact as a pool of $\text{CO}_{2(\text{aq})}$ is maintained at every time point. When the cultivation pH is 8 (Figure 4.10 C), the rate constant k^- is one order of magnitude smaller. This results in a significant accumulation of bicarbonate during the exponential growth phase, which is only slowly stripped after the cessation of the biological CO_2 -formation. The underestimation of the CO_2 -formation rate is significant in this case. These simulation findings show that an underestimation of the total CO_2 -production rate occurs whenever part of its dissolves in the medium.

Summarizing the above findings, we consider a system with three balanced pools with two system inflows and one outflow. The inflows are the biological CO_2 -production rate (flowing into $\text{CO}_{2(\text{aq})}$) and the CO_2 -inflow from the inlet air (flowing into the $p\text{CO}_2$ pool). The outflow from the system is the CO_2 in the outlet air. If the outflow of the system was the same as the inflow at any time during the cultivation, the calculated CO_2 production rate would be equal to the biological CO_2 -formation rate $\Pi_{\text{CO}_2, \text{tot}}$. However, during the cultivation the equilibrium between CO_2 and HCO_3^{-1} causes a netflow into HCO_3^{-1} , which acts as a temporary sink and thereby reduces the amount of carbon leaving the system via the outlet air during the exponential phase.

9.4.1 Experimental Illustration

The precise parameter values finally define whether the calculated CO_2 -formation rate is approximately equal to the true CO_2 -formation rate. Some parameters like the rate constants k^+ and k^- and the $k_L a$ -value depend on the used cultivation pH and the aeration regime, respectively. Every change favoring the mass transfer term relative to the hydration term ought to increase the instantaneous CO_2 -recovery. By contrast, the opposite effects should be observed by increasing the netflow towards

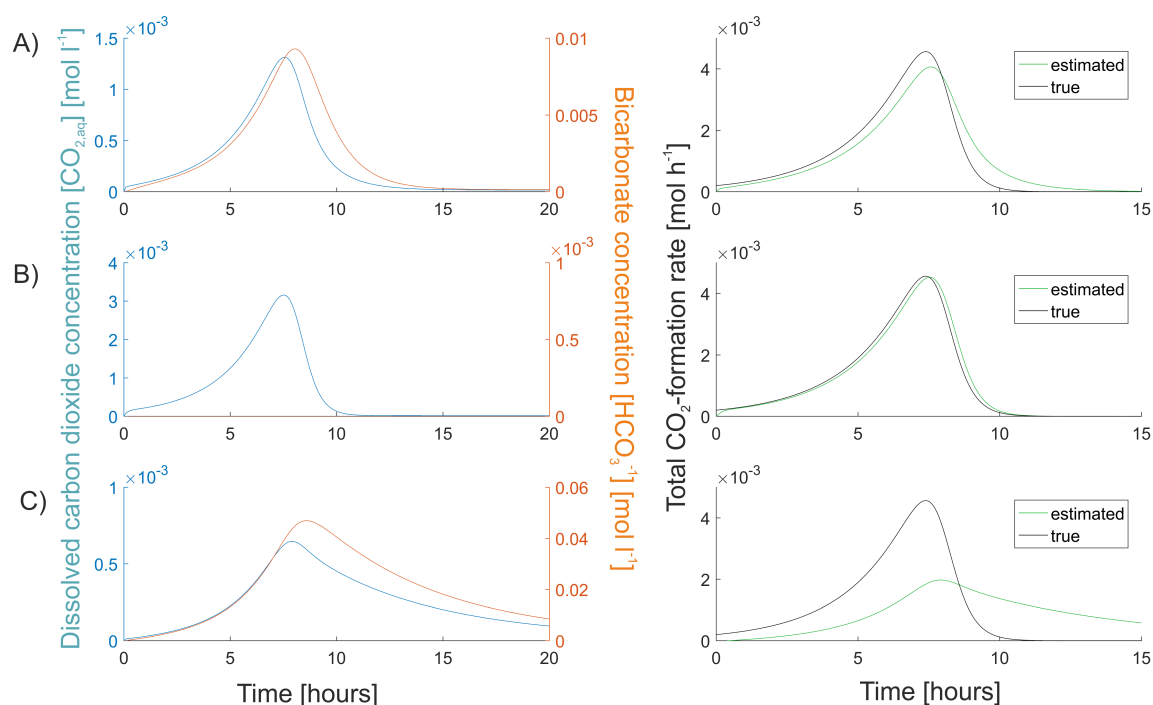


Figure 4.10: Simulated time course of $\text{CO}_{2(\text{aq})}$ and HCO_3^- as well as the true and estimated CO_2 -formation rates for three different scenarios of bicarbonate formation. A) bioreactor cultivation at pH=7, B) rate constant k^+ set to zero and C) bioreactor cultivation at pH=8.

bicarbonate which can be achieved by altering the pH value. To validate the correctness of the above made predictions four experiments under controlled bioreactor conditions were performed, varying the above mentioned parameter values.

- 1) Batch cultivation at pH=7, 1 vvm [volume of air per volume liquid per minute in l (l min)^{-1}]. This is considered as reference scenario. Under this condition, all extracellular phenotyping experiments of anaplerotic deletion mutants were performed.
- 2) Batch cultivation at pH=6.4, 1 vvm. A lower pH value ought to diminish bicarbonate formation and thus increase the relative importance of liquid-to-gas mass transfer.
- 3) Batch cultivation at pH=7, 3 vvm. Increasing the volumetric gas flow is supposed to favor liquid-to-gas mass transfer by maintaining a steeper concentration gradient of CO_2 between the liquid phase and gas phase. Blombach et al. [22] showed by direct measurement of CO_2 in the aqueous phase during a cultivation of *C. glutamicum* that this treatment kept the $\text{CO}_{2(\text{aq})}$ concentration below 10 mbar for biomass concentrations below 10 g l^{-1} , whereas it increased exponentially to approximately 100 mbar at a biomass concentration of 10 g l^{-1} for 0.1 vvm.
- 4) Batch cultivation at pH=8, 1 vvm. A higher pH value increases k^+ and thereby shifts the equilibrium between $\text{CO}_{2(\text{aq})}$ and HCO_3^- towards bicarbonate.

Figure 4.11 depicts the time course of the volume fraction of CO₂ in the exhaust gas stream during the different cultivation conditions. It is clearly visible that a prolonged CO₂-evolution after the onset of the stationary phase is observable during the cultivation at pH=8. This is apparently the result of an increased bicarbonate pool accumulated at this pH, which is slowly stripped out when the biological CO₂-production ceases. However, even 15 hours after reaching the stationary phase the presence of a significant bicarbonate pool size can be seen. At 23 hours after inoculation the pH was shifted from 8 to 7 by feeding HCl into the system, increasing the backward rate constant by one order of magnitude. With the increasing back-reaction towards CO_{2(aq)} a clear spike in the exhaust gas CO₂ concentration can be seen, proving the presence of a considerable bicarbonate pool at this time point. This is in agreement with the simulation results from above which predicted such a presence at pH=8. When the pH was shifted from pH=7 to 6.4 in the two cultivations performed at pH 7 at the same time point, no increase in exhaust gas CO₂ concentration was detectable, which suggest that the bicarbonate pool had already been completely stripped. This agrees with the simulation results from above.

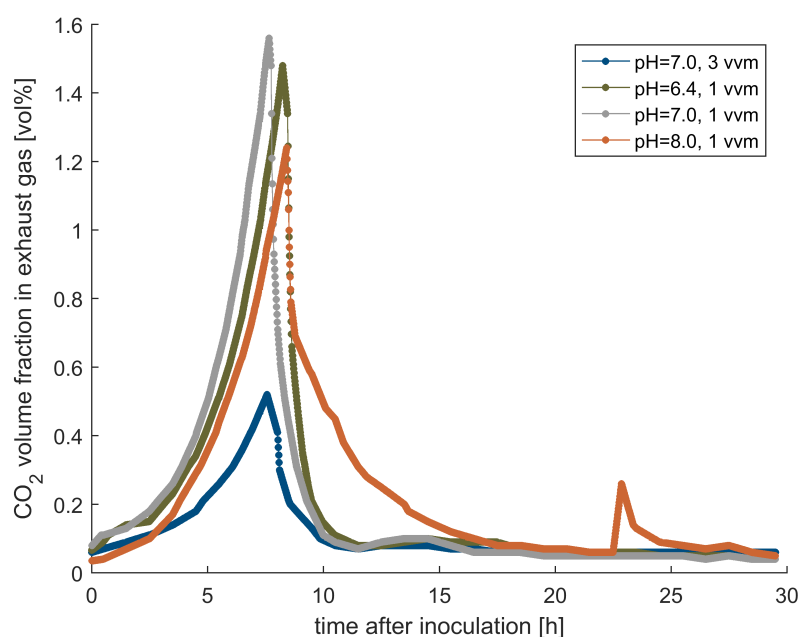


Figure 4.11: Time course of the volume percentage of CO₂ in the exhaust gas stream for four considered cultivation regimes. Each cultivation was performed identically to the cultivations of anaplerotic deletion mutants described in section 7.1 except for the change of aeration rate and pH control.

For each cultivation the carbon balance was calculated according to equation (4.5), the result of which is depicted in Figure 4.12. The actual closure crucially depends on the assumed carbon content in CDW. Here, the value from Buchholz et al. [24] was used which is 0.5 g carbon (g CDW)⁻¹. It can be seen that for the 3 vvm scenario an almost closed carbon balance is reached, whereas all other scenarios produced non-closed carbon balances, particularly the pH 8 scenario. None of the cultivations exhibited major side-product formation (data not shown).

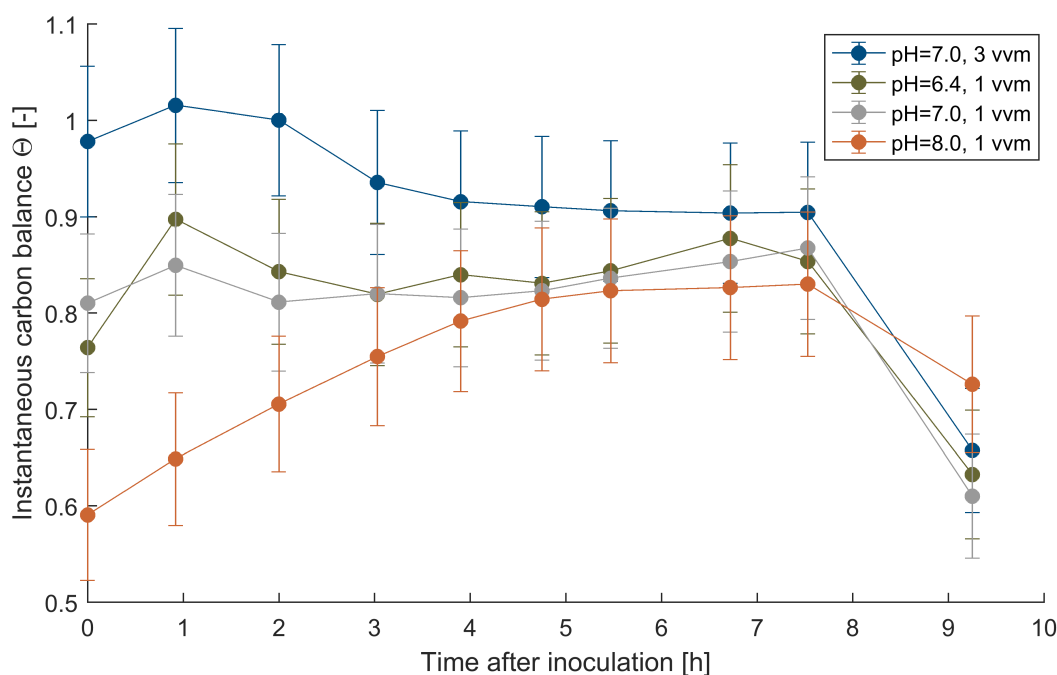


Figure 4.12: Time course of the instantaneous carbon balance for *C. glutamicum* WT under different pH and aeration conditions.

9.5 Summary

It could be shown that the proposed scheme for rate estimation yields consistent results across the set of anaplerotic deletion mutants. The obtained specific rates were validated through a carbon balancing approach. Thereby, it was demonstrated that the estimation of the CO_2 -formation rate is systematically biased in batch cultivations. Although the non-closure of carbon balances in batch cultivations was reported previously, the consequences for specific rate estimation had not been worked out as thoroughly as in this thesis. Furthermore, the derived specific extracellular rates allow to deduce hypotheses about the flux distribution in selected mutant strains and provide a benchmark for more detailed omics-phenotyping.

10 Metabolome analysis

10.1 Intracellular metabolome

Section 9.3 highlighted altered physiological states for some anaplerotic mutant strains. To gain further insight into the metabolism of each mutant, an intracellular metabolome analysis was performed. The samples for this analysis originated from a dedicated bioreactor cultivation of each deletion mutant, during which samples were drawn at two time points corresponding to a BV concentration of 5 and 10 $\mu\text{l ml}^{-1}$, respectively. These BV concentrations represent target values which determined the sampling time points. The actual BV concentration in each sample was measured after sampling and used to calculate the extraction volume. To ensure comparability, eight strains were cultivated in parallel and subjected to identical and isochronous sample processing in subsequent steps. Since the ΔPEPCx mutant showed no altered phenotype in the previous section, it was omitted from the set of strains to be analyzed, leaving eight strains. The samples of all eight strains were spiked with identical IS and were measured in one acquisition batch on the LC-ESI-QqQ MS system. The measurement was performed according to section 7.5 of the Material and Methods section. Organic acids and unstable sugar phosphates were measured within 24 h from the extraction of samples. The tight coupling of sample generation and measurement proved crucial to preserve signal on mass traces corresponding to unstable metabolites such as NADH, GTP, GDP, ACCOA. It is well established that these metabolites readily and quickly degrade [164]. In summary, this procedure ensured that no artifacts arise in following comparative metabolome analysis and that the concentration measurements on unstable metabolites could be performed.

Figure 4.13 shows the calculated intracellular concentrations of GLX, PEP, PYR, ALA, VAL, GLY, SER and ACCOA. For the statistical evaluation an eight-way ANOVA was performed to determine if the concentration of one metabolite was significantly altered among the investigated strains. If so, a Tukey-Kramer post-hoc test for significant pairwise differences between group means was performed using a significance level of 0.05. From Figure 4.13 it becomes apparent that the GLX shunt is indeed active in the $\Delta(\text{PCx PEPCx})$ strain for its intermediate GLX could be detected. This observation corroborates the hypothesis suggested by the extracellular rate measurements discussed in section 9.3. The GLX shunt acts as anaplerotic reaction in this strain due to the absence of PCx and PEPCx activity. In the absence of C3-carboxylation activity at the anaplerotic node, accumulation of the substrate pools PYR and PEP can be expected. Indeed, PEP is significantly higher concentrated in the $\Delta(\text{PCx PEPCx})$ strain in comparison to all other strains. PYR, on the other hand, does not discriminate strains. Concentration measurements on ALA and VAL, however, do allow a clear distinction of the $\Delta(\text{PCx PEPCx})$ mutant. Since these two metabolites are directly derived from PYR, it seems plausible that the lack of statistical significance of strain differences in their precursor metabolite PYR is most likely due to high technical errors in its quantification rather than a true absence of difference. It is well established that the accurate quantification of organic acids in cell extracts represents a veritable challenge [209]. The metabolite pool of GLY turns out to exhibit one of the most significant concen-

tration changes, clearly distinguishing the $\Delta(\text{PCx PEPCx})$ mutant. GLY, in turn, is derived from SER, the concentration of which is also increased in the $\Delta(\text{PCx PEPCx})$ mutant. Contrary to GLY, however, SER concentration also appears to be equally increased in both ΔPEPCk deletion strains.

Summarizing, it appears that the missing carboxylation rate cannot be matched by the PDHC activity in the $\Delta(\text{PCx PEPCx})$ mutant for substrate pools like PEP as well as amino acids derived from the lower glycolytic intermediates appear to accumulate intracellularly. The detection of the intermediate GLX of the GLX shunt clearly proves the activity of this pathway in the $\Delta(\text{PCx PEPCx})$ mutant. Wendisch et al. [190] was the first to suggest that the carbon-source dependent regulation of this pathway is mediated by intracellular ACCOA concentrations. However, as seen in Figure 4.13 it becomes apparent that the intracellular concentrations did not vary significantly with respect to strain background. Not even the mean values suggest a correlation with GLX shunt activity. Therefore, this often cited hypothesis [182, 172] could not be validated with the made intracellular measurements.

Apart from the metabolites from above, NADPH and NADH were quantified alongside their oxidized analogs. No significant difference was detected in the reduced forms of these co-factors. However, since the mean of relative standard deviation over all mutants for NADH and NADPH concentrations amounts to 46 % and 48 %, respectively, the quantification can hardly be called successful. This high technical error, impeding a precise quantification, cannot be traced back to inaccuracies in pipetting or BV concentration measurements since these factors also apply to all other metabolite quantifications in the same sample. Since the mean relative standard deviation for the calculated intracellular concentration of HOM, for instance, amounts to 9 %, these factors appear to have been controlled quite well. The most likely reason for the observed coefficients of variation is metabolite instability. These redox equivalents are known to be quite sensitive to oxidation and degradation [164]. Slightly different temperature time courses, residual enzymatic activity in metabolite extracts and oxidation most likely account for the observed differences. Therefore, this experiment did not allow accurate conclusions about the redox state in each mutant.

Figure 4.14 shows concentration measurements for MAL, αKG , GLU, PRO, HOM and ASP. For MAL and αKG the difficulty of precise intracellular concentration measurements of organic acids becomes apparent. Although mean concentration values of MAL seem to be increased in all mutants deficient in either ME or ODx activity, none of the changes is significant. The ASP metabolite pool is most closely correlated with the growth rate. The lowest concentrations are observed for all mutants with reduced growth rate. Apparently, its supply appears to be limiting the growth as the restored growth rate of the $\Delta(\text{PEPCk ME})$ mutant in comparison to the single deletion strain ΔPEPCk goes hand in hand with increased ASP supply. The concentration pattern of ASP seems to reflect itself in the HOM pool, which is derived from the former through three intermediate reaction steps, consuming two NAD(P)H molecules and one ATP molecule². In contradistinction to ASP, however, HOM is clearly higher concentrated in the ΔPEPCk mutant. Since the reaction sequence from ASP to HOM is redox-

²The biosynthesis of HOM from ASP in KEGG PATHWAY map cgb00260 proceeds through cg0306-cg0307-cg1337. The products of cg1337 Homoserine dehydrogenase and cg0307 Aspartate semialdehyde dehydrogenase catalyze NADPH dependent reactions.

dependent, one may be tempted to attribute this observation to redox balancing. The biosynthesis of GLU from α KG³ and of PRO from GLU are also redox-dependent⁴. Conspicuously, both pools are higher concentrated in the Δ PEPCk mutant. However, this hypothesis does not hold up generally. SER concentrations are higher in the Δ PEPCk mutant, although SER biosynthesis produces NADH⁵. Furthermore, the biosynthesis of VAL from PYR is NADPH-dependent but VAL does not show an increase in this mutant.

In summary, many significant changes were observed across the set of deletion mutants for a multitude of metabolite pools. However, these changes cannot be interpreted in light of the co-factor dependence of central carbon metabolism reactions. The fact that the removal of a redox-dependent enzyme like ME restores the growth rate of the Δ PEPCk mutant suggests, nonetheless, an involvement of the redox balance.

³The biosynthesis of GLU from α KG in *C. glutamicum* was shown to be catalyzed by two enzymes [173, 46]: Glutamine 2-oxoglutarate aminotransferase (GOGAT), the two subunits of which are encoded by cg0229 and cg0230, and Glutamate dehydrogenase (GDH), encoded by cg2280. Both enzymes require NAD(P)H as cofactor.

⁴KEGG PATHWAY cgb00330 lists two pathways for PRO biosynthesis in *C. glutamicum*: The first is encoded by cg2588-cg2586-cg0490, the second by cg0129-cg0490. The gene product of the latter gene of both pathways, cg0490, Pyrroline-5-carboxylate reductase, requires NAD(P)H.

⁵Biosynthesis of SER from 3PG is catalyzed by the enzymes encoded by cg1451, cg0948 and cg2779. The gene product of cg1451, Phosphoglycerate dehydrogenase, catalyzes the NAD⁺-dependent step (KEGG PATHWAY map cgb01230).

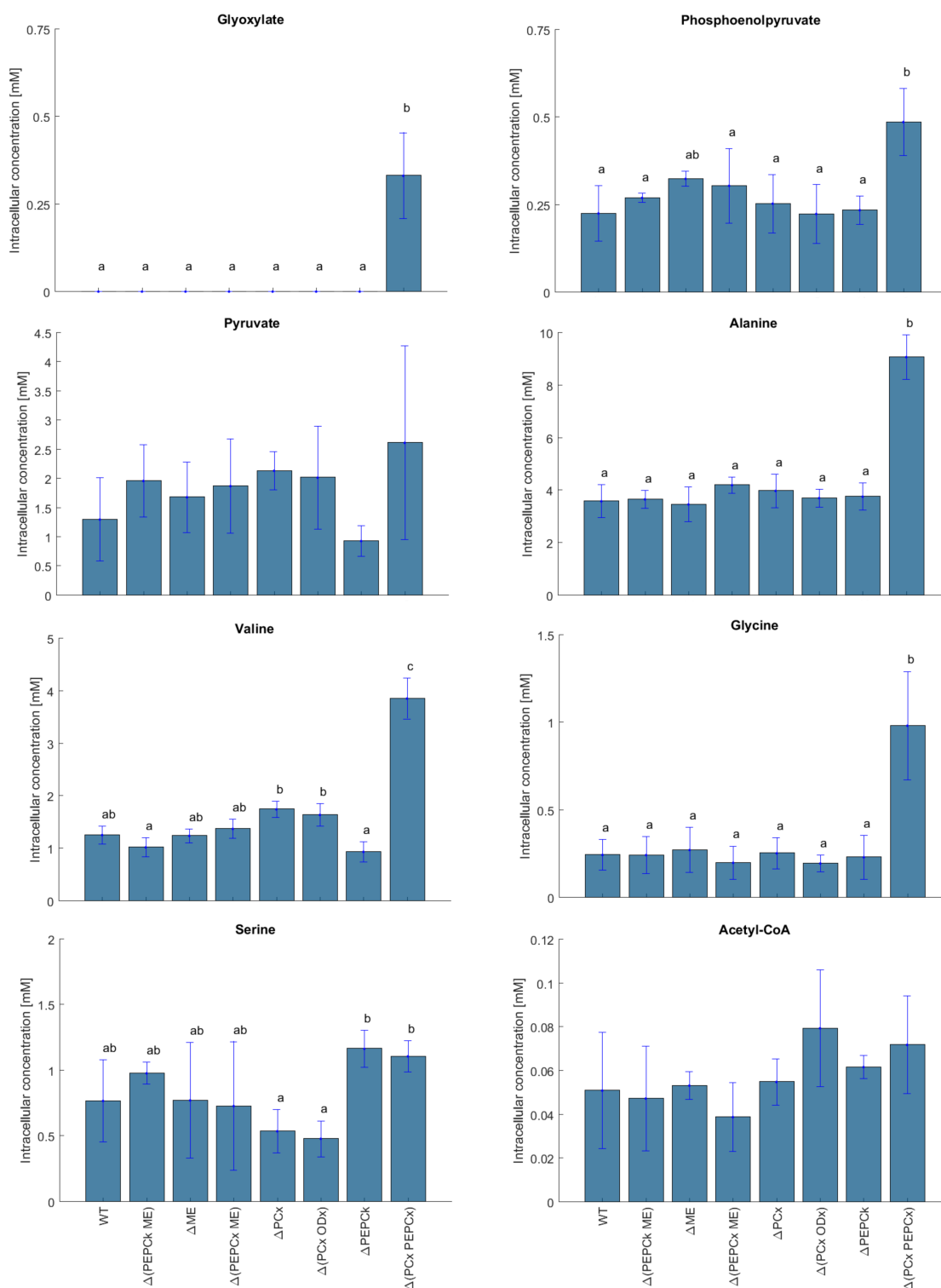


Figure 4.13: Intracellular concentrations of selected organic acids, amino acids and ACCoA in anaplerotic deletion mutants during growth in CGXII medium with glucose as sole carbon source. If no bar is annotated with a superscript letter, the genetic background had no significant effect on the shown concentrations. If one is present, though, the group means of deletion mutants having a different superscript letter are statistically different. If the letter is identical for any two mutants, on the other hand, their group means are not significantly different.

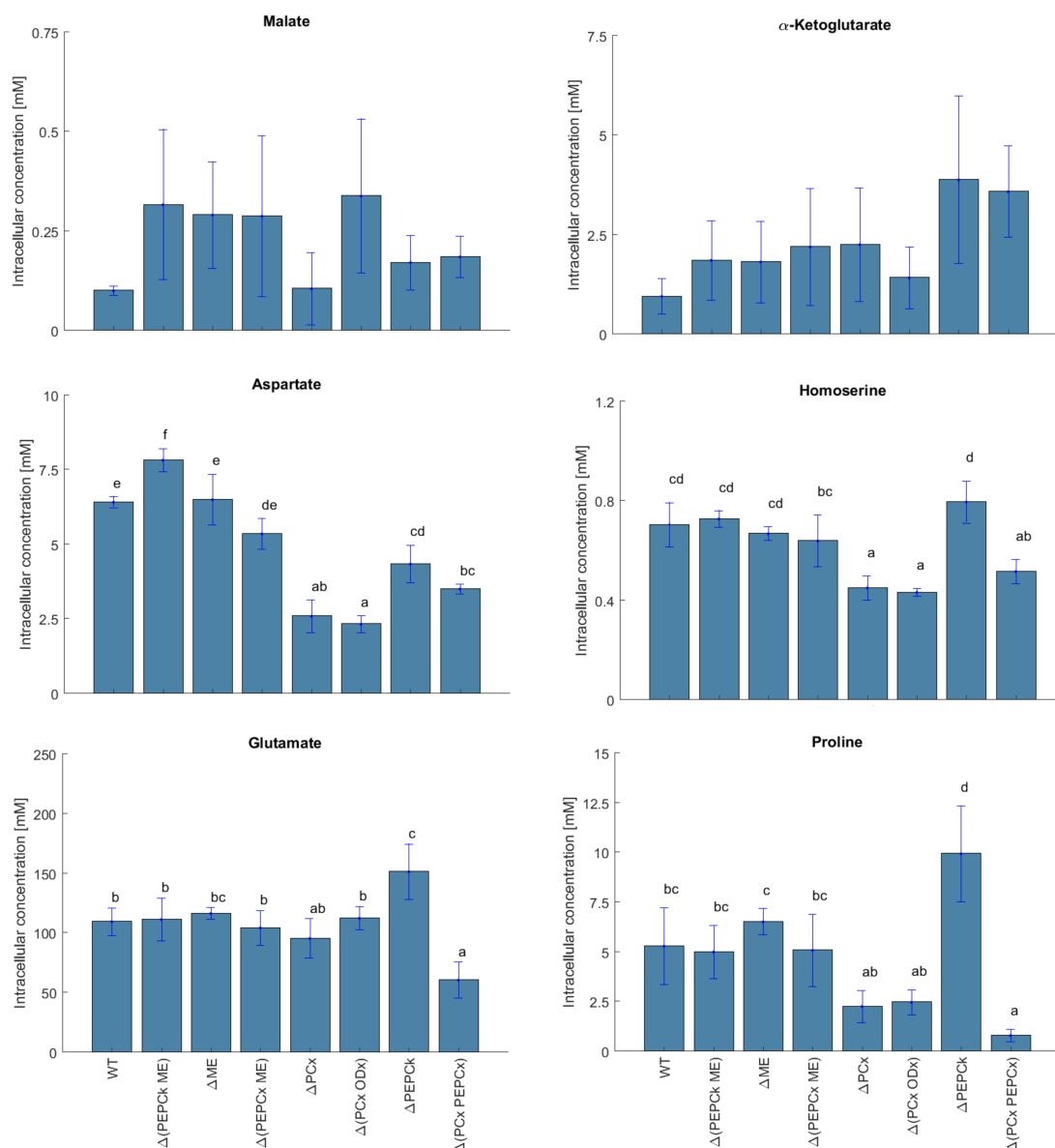


Figure 4.14: Intracellular concentrations of selected organic acids and amino acids in anaplerotic deletion mutants during growth in CGXII medium with glucose as sole carbon source. If no bar is annotated with a superscript letter, the genetic background had no significant effect on the shown concentrations. If one is present, though, the group means of deletion mutants having a different superscript letter are statistically different. If the letter is identical for any two mutants, on the other hand, their group means are not significantly different.

10.2 Exometabolome

The above highlighted intracellular concentration changes may lead to a different exometabolome profile in some strains, which depending on the magnitude may have consequences for ^{13}C -MFA. The samples for the exometabolome analysis were generated and analyzed in the scope of the above mentioned metabolome experiment. Therefore, exometabolome samples were generated at two time points. For each time point the biomass-specific amount of each metabolite was calculated by normalizing its total amount to the biomass at this time point. Figure 4.15 shows the specific amounts of selected metabolites for the two time points. Apparently, some of the metabolite pools that accumulated intracellularly are excreted into the extracellular space. This is the case for GLY, GLX, and PEP in the $\Delta(\text{PCx PEPCx})$ mutant strain as well as for GLU and GLN in the cultivation of the ΔPEPCk mutant. The increased extracellular concentration in the extracellular domain mirrors the intracellular concentration in the respective mutant strains. Since the extracellular concentration are in lower μM range, whereas the calculated intracellular concentrations are in the mM range, it seems plausible that the higher extracellular concentrations are driven by a concentration gradient.

The PYR metabolite pool, on the other hand, shows only a slight increase in intracellular concentrations in the $\Delta(\text{PCx PEPCx})$ mutant, but a substantial increase in extracellular concentrations. When the intracellular concentration of 2.61 mM is converted to a CDW-specific amount, a value of $8.44 \mu\text{mol (g CDW)}^{-1}$ is obtained. Strikingly, based on the data in Figure 4.15, the specific extracellular amount reaches approximately six times the intracellular amount. Since other strains deficient in PCx activity show intracellular concentrations that are statistically not different from the one in the $\Delta(\text{PCx PEPCx})$ strain, a diffusion-driven mechanism can be excluded. The identified genetic alterations from Table 4.4 show a mutation in the *mctC* gene (cg0953) in the $\Delta(\text{PCx PEPCx})$ strain. This gene has been shown to be essential in *C. glutamicum* for the uptake of PYR in the study of Jolkver et al. [85]. Based on this, it can be hypothesized that the genetic alteration in this transporter may lead to an enhanced export of PYR into the medium. The reason why this may be a beneficial or essential adaptation in the $\Delta(\text{PCx PEPCx})$ strain remains to be shown.

The remaining set of measured sugar phosphates, organic acids and amino acids in this targeted analysis does not show significant differences with respect to the genetic background. The determined extracellular concentrations of the total set of measured metabolites at the second time point correspond to a total specific amount of $274 \mu\text{mol carbon (g CDW)}^{-1}$ for the WT. The exometabolome of the $\Delta(\text{PCx PEPCx})$ strain corresponds to $786 \mu\text{mol carbon (g CDW)}^{-1}$, which gives a clear indication of a more extended overflow metabolism. Furthermore, during the cultivation of the ΔPEPCk strain larger quantities of metabolites also accumulated extracellularly, amounting to $657 \mu\text{mol carbon (g CDW)}^{-1}$. This is, however, primarily due to the excretion of GLU and GLN.

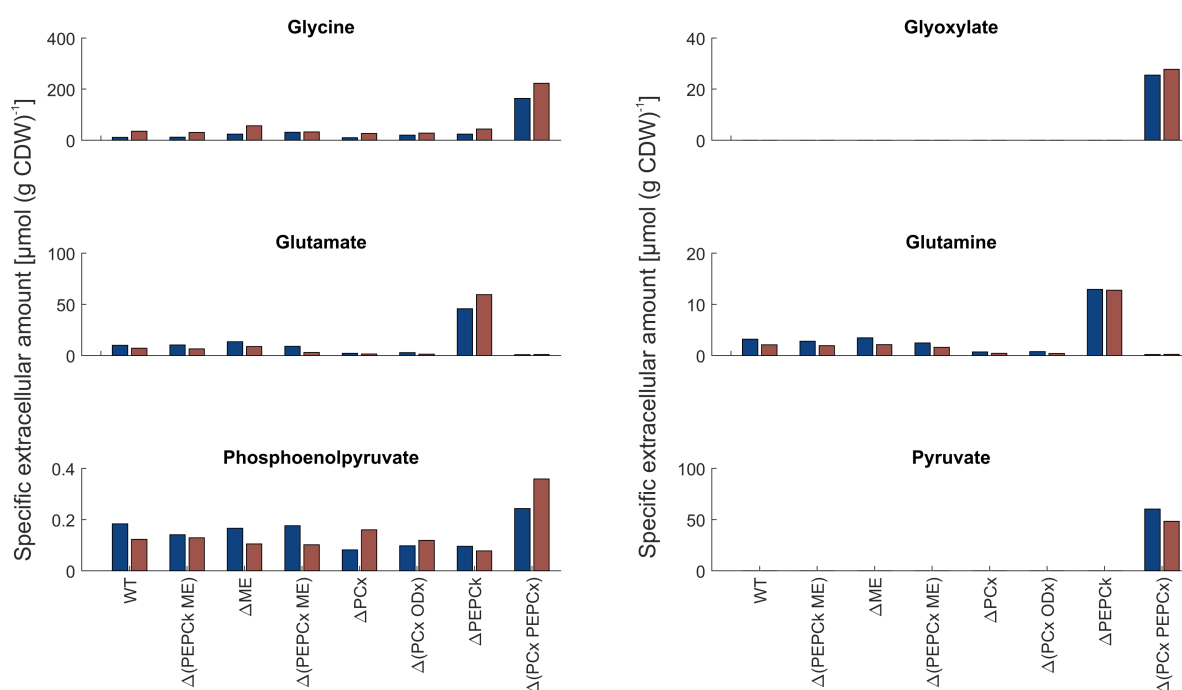


Figure 4.15: Specific extracellular concentrations of selected amino acids and organic acids for anaplerotic deletion mutants cultivated in controlled bioreactor cultivations. Blue bars represent the specific extracellular concentration at the first sampling time point, whereas orange bars represent those at the second sampling point of the metabolome experiment outlined above.

11 Proteome analysis

11.1 Data generation and data handling

To gain a detailed insight into the distinct adaptation of each mutant, a proteome profiling experiment was performed. The samples for this analysis originated from the metabolome experiment outlined above and were drawn after all quenching samples for the metabolome analysis had been taken. Shortly, from each reactor 10 ml culture broth was sampled in quintuplicate by centrifugation at 4500 rpm for 10 minutes (GS-15R Centrifuge from Table 3.4). After the supernatant was decanted, the samples were immediately frozen at -20°C in metal racks. The sample processing followed the procedure described in Voges and Noack [183] in every aspect. All five samples per strain were independently lysed and digested and will therefore be denoted 'digest replicate' in the following. Each digest replicate, in turn, was injected three times into the LC-MS/MS system, henceforth referred to as 'injection replicate'. Thus, for each of the eight strains 15 injections were available, totaling 120 LC-MS/MS acquisitions for all eight strains, which were acquired in one batch.

The targeted data extraction pipeline described in subsection 7.7 of the Materials and Methods section yielded a list of integrated fragment ion mass traces corresponding to the light and heavy isotopes of at most 12 fragment ions for each included peptide. For each protein the fragment mass traces of at most ten peptides were extracted in a targeted manner. The peptides were selected based

on the corresponding precursor ion intensity in the survey scan that was used to build the ion library for data extraction. Thus, for each protein optimally 120 features were observed across all samples. The real number of mass transitions per protein may be lower than that since peptides with a FDR higher than 0.1 % were not exported. The total dataset comprises 145,586 rows corresponding to light and heavy isotopes of 72,793 fragments and 120 columns corresponding to injections. The complexity of the dataset at hand requires dedicated software not only for efficient data handling, but also to harvest the immense information content hidden in the extensive measurements.

Given the limitations of the manufacturer software package (PeakView, MarkerView) with respect to advanced statistical modeling and fragment ion selection, a MATLAB-based data storage framework was established in cooperation with Mathias Papenfuß, MSc Biotechnology RWTH Aachen, based on the MATLAB *struct* data type. The implemented approach harvests the hierarchical organization of proteome data from MS/MS instruments which is depicted in Figure 4.16. Based on this data storage format, a comprehensive data processing pipeline was implemented that is presented in the following.

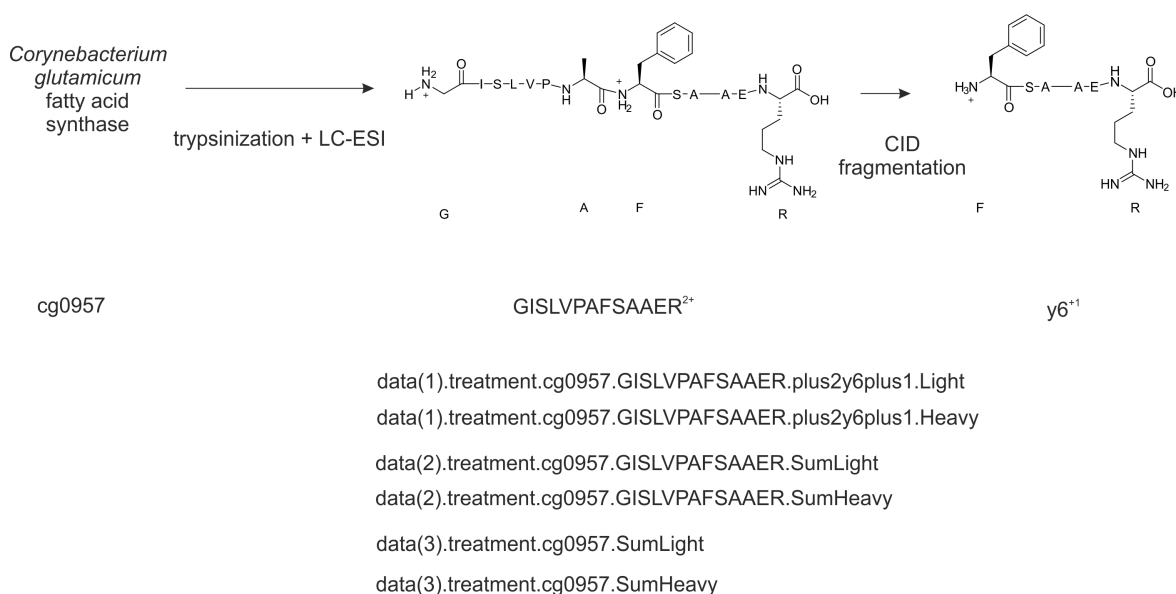


Figure 4.16: Exemplary data organization for the product of ORF cg0957 of *C. glutamicum*. Here 'treatment' denotes the physiological state of interest. The protein name forms the next level of the structure. The field names of the protein substructure correspond to the sequences of measured peptides of the protein. The measured mass transitions of each peptide are fields of the peptide structure. The fields of the lowest level substructure contain the mass trace intensities corresponding to the light and heavy isotope of the respective fragment ion.

11.2 Consistency and data quality check

As a first check of the data quality, the sum of all ion trace intensities corresponding to the ¹⁵N-labeled isotopes of fragment ions pertaining to one protein is computed. This summation yields one abun-

dance feature per protein in each sample. Since each sample contains the same amount of identical ^{15}N -labeled IS, the summed intensity of the ion traces corresponding to the ^{15}N -labeled isotope of each fragment should ideally be identical. Surely, the ^{15}N -labeled IS is employed to compensate for sample matrix dependent ion suppression and experimental variations, such that the summed peak areas are most likely not equal across samples. Rather, they can be anticipated to vary according to the strength of ion suppression and efficiency of the trypsinization in each sample. Notwithstanding, in the absence of systematic influences on the ion trace intensities, the variance in summed ^{15}N mass trace intensities across all samples should be stochastic and therefore exhibit no structure, i.e. no clustering. To detect a possible clustering, a principle component analysis (PCA) was performed on the ^{15}N fragment ion mass trace intensities summed per protein. In this setting proteins are variables, whereas each MS acquisition represents one observation. In essence, PCA finds a synthetic new variable as linear combination of the original variables, along which the variance across all samples is maximized.

The resulting score plot in Figure 4.17 shows that indeed no systematic clustering of the summed ^{15}N fragment ion trace intensities is apparent such that the data quality is not in doubt. The summed ^{14}N -ion trace intensities, on the other hand, may reveal a clustering since the amount of unlabeled protein in each sample stems from distinct conditions. A differential gene expression of one protein between these conditions should reflect itself in increased variance of the summed ^{14}N fragment ion intensities for this protein across samples in comparison to a non-differentially expressed protein. Indeed, performing the same analysis of the summed mass trace intensities corresponding to all unlabeled fragment ions of a protein reveals that all samples from the $\Delta(\text{PCx PEPCx})$ strain separate from those of other strains along the second principle component (see Figure 4.18). Since the samples cluster according to the genetic background and not according to some hidden variable, the data quality can be regarded as overall good. Furthermore, a clustering tendency for the ΔPCx and $\Delta(\text{PCx ODx})$ strains is apparent since all samples pertaining to these strains cluster in the upper half of the lower cluster. This raises the expectation that their expression profile is more similar to each other than it is to other strains.

The loading plot for the second principle component of the PCA analysis on summed ^{14}N ion intensities is shown in Figure 4.19. The loading of each variable on one principle component essentially measures the contribution of the corresponding variable to the principle component in question. Since the principle component, in turn, indicates the direction of maximum variance, high loadings on one of the first principle components indicate proteins that contribute the most to the observed variance across all samples. If the variance mainly stems from differential gene expression, then the loadings indicate differentially expressed proteins. The loading plot for the second principle component in Figure 4.19 identifies the protein products of ORFs cg0759, cg0949, cg1737, cg3047, cg3048, cg2559, and cg2560 as proteins contributing the most to the second principle component. With the argument from above, these proteins are most likely to be differentially expressed between samples of the upper and lower cluster of Figure 4.18.

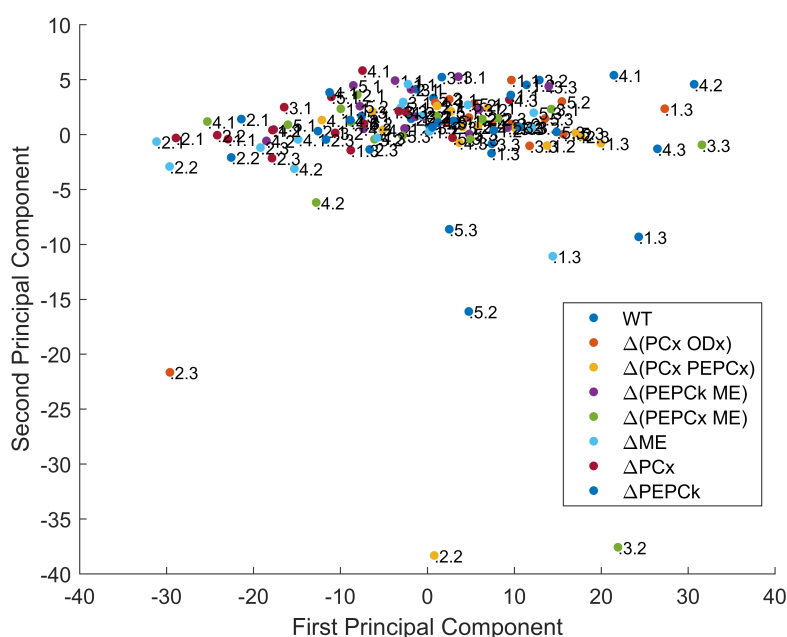


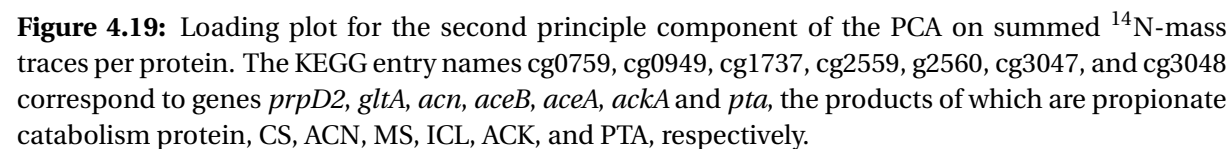
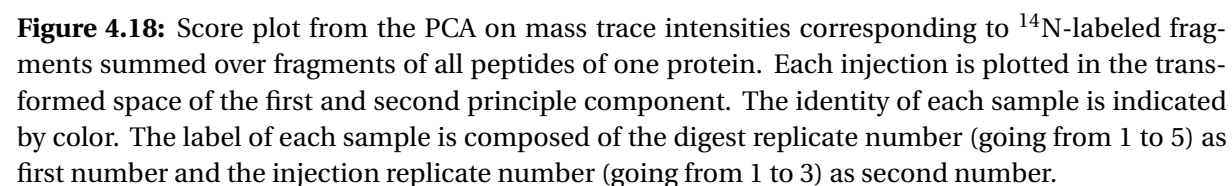
Figure 4.17: Score plot from the PCA on mass trace intensities corresponding to ^{15}N -labeled fragments summed over fragments of all peptides of one protein. Each injection is plotted in the transformed space of the first and second principle component. The identity of each sample is indicated by color. The label of each sample is composed of the digest replicate number (going from 1 to 5) as first number and the injection replicate number (going from 1 to 3) as second number.

The gene products of these ORFs correspond to the propionate catabolism protein, CS, ACN, ACK, PTA, MS, and ICL, respectively. The latter enzyme is the initial enzyme of the GLX shunt, the activity of which in the $\Delta(\text{PCx PEPCx})$ mutant could be proven in the previous section. This finding already indicates that an adaptation of the proteome profile has taken place in this mutant alongside the SNP mutations given in Table 4.5. When metabolism is based on anaplerotic reactions involving C3-carboxylation (PCx and PEPCx), the TCA cycle activity is only needed for oxidative decarboxylation, substrate level phosphorylation and provision of precursors for the glutamate family of amino acids. When anaplerotic activity is catalyzed by the GLX shunt, on the other hand, the initial reactions of the TCS cycle, i.e. CS and ACN need to account for all of the above as well as the anaplerotic carbon flux. This adaptation is clearly seen in the obtained loadings since both ICS and ACN have high loadings for the second principle component.

11.3 Fold change estimation using Poisson regression

11.3.1 Motivation

The above presented clustering approach is rather insensitive with respect to small expression changes between strains and does not harvest the additional measurements performed on the ^{15}N -labeled isotopes of each fragment ion. Therefore, this section is devoted to the development of a statistical framework to estimate precise fold changes (FCs) between an arbitrary number of conditions.



In particular, an extension of the regression approaches outlined in the Introduction to the case of metabolic labeling is presented, which allows the estimation of isotope ratios in individual digest replicates from which the FC between treatments is estimated across digest replicates. This strategy overcomes the disadvantage of metabolic labeling strategies that allow for only two plexes in one sample, i.e. light and heavy. Moreover, isotope ratios are ordinarily computed for each peptide of a protein since peptides are the measured analytes [183]. It is an open question how to integrate the isotope ratios of different peptides of a protein to arrive at protein-level conclusions. A regression approach on the other hand allows to reconcile all data in one estimation. In the following the framework is presented by sequentially introducing approaches to estimate:

1. an isotope ratio in one digest replicate.
2. an isotope ratio over several digest replicates from the same condition (same isotope ratio).
3. a FC between two conditions from digest replicates stemming from different conditions.

11.3.2 Estimation of isotope ratios in one digest replicate

A log-transformed peak area from injection replicate z from a sample containing light and heavy isotopes of peptides can be modeled as:

$$y_{i,p,f(p),z} = \tau + I_i + \Pi_p + \Phi_{f(p)} + \epsilon_{i,p,f(p),z} \quad (4.17)$$

$y_{i,p,f(p),z}$ denotes the log-transformed peak area which shall be differentiated from the raw mass trace intensity $Y_{i,p,f(p),z}$ by its lowercase letter. The indices indicate that the mass trace intensity stems from i^{th} isotope of fragment $f(p)$ of peptide p from a given protein. τ is the overall mean of the vector $\mathbf{y}$ of all data points. I_i is the effect of the i^{th} isotope as additive deviation from the overall mean τ . The factor 'isotope' has always two levels in metabolic labeling, i.e. light (^{14}N) and heavy (for instance ^{15}N). Π_p is the effect of the p^{th} peptide since digest yields and ionization efficiencies can be expected to be different for peptides of one protein. $\Phi_{f(p)}$ is the effect of the f^{th} fragment of peptide p . The fact that factor 'fragment' (Φ) is nested within 'peptide' (Π) is accounted for by the nested index $f(p)$. Appendix 7 exemplarily shows the Design Matrix for the regression problem of equation (4.17) for the case where the factors 'isotope', 'peptide' and 'fragment' have two levels each.

As a basic fact of linear regression, only $(n - 1)$ coefficients are identifiable for the n levels of each factor. A constraint is usually added requiring that all coefficients for the levels of each factor add to zero [92]. In this case τ has the interpretation of the overall mean and each coefficient models the deviation of the values corresponding to one level of a certain factor from the overall mean. Alternatively, the so-called 'reference' coding in linear regression takes one specific level of each factor as zero and just models the additive change of all other levels of a particular factor relative to the factors' reference level. When logged mass trace intensities are used, the additive change corresponds to a

multiplicative effect of the respect level on the original scale. Hence, when the level 'heavy' is taken as reference for the factor 'isotope' in equation (4.17), the exponential of the coefficient for the level 'light' directly represents the L/H-ratio.

11.3.3 Estimation of isotope ratios over several digest replicates

To estimate a L/H ratio over several digest replicates that ought to yield the same isotope ratio, one may simply add these data points with their grouping (isotope, peptide, fragment) to the regression problem, providing replication with each group. However, this may not be appropriate when matrix-dependent ion suppression effects occur or the enzymatic digest efficiencies vary between the replicates. Both effects introduce a non-random deviation between peak areas $y_{z,i,p,f(p)}$ of the same group and reflect themselves in different terms in the regression problem. Differing tryptic digest efficiencies can be expected to affect all peptides of a sample in the same way, reducing all peak areas belonging to the group identified by the respective replicate. This can be modeled as one main effect per digest replicate, accounting for one constant deviation from the reference group mean over all replicates, which is constant for all peptides and both isotopes alike of one replicate. The problem can be stated as follows:

$$y_{i,p,f(p),r,z} = \tau + I_i + \Pi_p + \Phi_{f(p)} + R_r + \epsilon_{i,p,f,r,z} \quad (4.18)$$

The ion suppression effect depends on the amount of ion-suppression agents eluting concurrently to the affected peptide. Therefore, the ion suppression effect can be expected to manifest itself on the peptide level in specific replicates. This can be modeled as an interaction effect between replicate and peptide according to:

$$y_{i,p,f(p),r,z} = \tau + I_i + \Pi_p + \Phi_{f(p)} + R_r + \Pi_p * R_r + \epsilon_{i,p,f,r,z} \quad (4.19)$$

The asterisk in equation (4.19) denotes an interaction effect. When we have P peptides and R replicates, we have $(R - 1) * (P - 1)$ variables for the interaction effect. The approach to model digest replicates as separate factor is essentially equivalent to the scheme proposed by Clough et al. [31] for the evaluation of human epidemiological data. The latter approach models individual study subjects as separate factor which in the present case would correspond to digest replicates.

11.3.4 Estimation of fold changes from isotope ratios from several digest replicates

When metabolic labeling is used, only one FC can be estimated from one digest replicate since only two plexes are available. When several conditions are to be compared, they can be mixed with identical IS. Then, each condition can be quantified relative to the IS. Since each sample bears identical IS, the FCs of each condition to the IS can be used to compute FC between them. Therefore, to esti-

mate FCs between an arbitrary number of samples with respect to one condition, the isotope ratios in several digest replicates must be related to each other.

The developed procedure is essentially a synthesis of the modeling approaches from the studies mentioned in the Introduction: On the one hand, Hill et al. [74] proposed to estimate the relative isotopic abundance in one sample (since FCs between several conditions can be derived from a single sample using iTRAQ-labeling). Clough et al. [31], on the other hand, proposed an ANOVA model for label-free LC-MS proteomics data, modeling intersubject variability as factor nested within treatment (disease) group. The following procedure combines both approaches for application to metabolic labeling, in which relative isotopic abundance has to be estimated over several samples to arrive at FCs between conditions.

To estimate FCs one can simply introduce an additional factor 'treatment' to equation (4.19), coding one treatment or condition as reference. However, in each sample only the ^{14}N -areas relative to the ^{15}N -areas should vary according to treatments. This property can be coded by an 'isotope * treatment' interaction effect. With 'heavy' and the condition to which all others should be quantified set as reference levels, the coefficient of the interaction effect of treatment t with isotope 'light' represents the FC of treatment t with respect to the reference treatment. The factor 'replicate' denoting digest replicates becomes nested within 'treatment' since replicate 1 from treatment 1 is different from replicate 1 from treatment 2. We have $T * (R - 1)$ variables for replicates when T is the total number of treatments and R is the constant number of replicates per treatment. In this case, the interaction effect between replicate and peptide has $T * (R - 1) * (P - 1)$ variables. The problem then becomes:

$$y_{i,t,p,f,r,z} = \tau + I_i + T_t + I_i * T_t + \Pi_p + \Phi_{f(p)} + R_{r(t)} + \Pi_p * R_{r(t)} + \epsilon_{i,t,p,f,r,z} \quad (4.20)$$

Equation (4.20) models a digest replicate as separate effect $R_{r(t)}$. Within-group replication can be performed by repeatedly injecting the same digest replicate $r(t)$, providing in-group replicates z .

11.3.5 Parameter estimation and statistical analysis

The regression problem for all observations $\mathbf{y}$ can be more generally formulated according to equation (4.21):

$$\mathbf{y} = \mathbf{X} \cdot \boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (4.21)$$

$\mathbf{y}$ is the standing vector of all log-transformed integrated mass trace intensities, whereas $\boldsymbol{\beta}$ is the vector containing all coefficients for main effects and interaction effects, i.e. $I_2, T_2, \dots, \Pi_2, \Pi_3, \dots, \Pi_p, \Phi_{2(1)}, \dots, R_{R(T)}$. The matrix $\mathbf{X}$ represents the design matrix, which harbors the dummy variables for all included coefficients. These dummy variables take the value of 1 for all observations belonging to that particular level or combination of levels to which they belong and 0 otherwise. Hence, equation (4.20) represents one row of matrix equation (4.21).

The observations at hand are integrated mass trace intensities which are thus essentially a weighted sum of ion counts. Counts, in turn, are most appropriately modeled as a Poisson process. Since every weighted sum of integrated Poisson distributed variables is again Poisson distributed, fragment ion trace intensities can be assumed to be a realization of Poisson random variable. In particular, a Poisson regression model considers each observation Y_j to be a draw of Poisson random variable with scale parameter λ_j , i.e. $Y_j \sim \text{Pois}(\lambda_j)$, which is dependent on the linear regressors according to equation (4.22) according to:

$$\lambda_j = \exp(\mathbf{X}_j \cdot \boldsymbol{\beta}) \quad (4.22)$$

where $\mathbf{X}_j$ is the line of the Design Matrix corresponding to observation j . Maximum likelihood estimation is used estimate $\boldsymbol{\beta}$ from the data. In particular, the estimate $\hat{\boldsymbol{\beta}}$ is the root of the derivatives of the Poisson log likelihood function with respect each element of $\boldsymbol{\beta}$. These equations have no closed form solution and therefore $\boldsymbol{\beta}$ is estimated using a numerical scheme called iteratively reweighted least squares. Under regularity conditions this algorithm converges to the maximum likelihood solution. Given the estimate $\hat{\boldsymbol{\beta}}$, the fitted rates $\hat{\lambda}$ can be calculated according to equation (4.22). It can be shown that asymptotically the covariance matrix of parameter estimates $\hat{\boldsymbol{\beta}}$ of the Poisson regression can be estimated from equation (4.23):

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \cdot \text{diag}(\hat{\lambda}) \cdot \mathbf{X})^{-1} \quad (4.23)$$

Equation (4.23) holds when the basic property of Poisson variable, i.e. the equality between expected value and variance is fulfilled. It has to hold $E(Y_j) = \lambda_j = \text{Var}(Y_j)$. This would be the case if the variance in observed integrated mass trace intensities within one group originated only from the ion counting process itself. However, even for repeated injections of one digest replicate the fragment ion mass trace intensities likely differ as a result of other factors. Loss of post-translational modifications during autosampler storage [183] and inaccuracies during integration of fragment ion mass trace signals will contribute to the variance within one group. Thus, the actual Poisson process (the MS acquisition) is overlaid by additional processes. These processes add additional variance to the integrated fragment mass trace intensities obtained after SWATH data processing. The deviation of the real variance from the theoretic value of a Poisson process can be expressed as 'overdispersion parameter' which ought to have the value of one to uphold the assumption of a Poisson process. In the Poisson regression overdispersion is accounted for by estimation of the overdispersion parameter ϕ from the data according to the equation (4.24):

$$\hat{\phi} = \frac{\sum_{j=1}^J \frac{(y_j - \exp(\mathbf{X}_j \cdot \hat{\boldsymbol{\beta}}))^2}{\exp(\mathbf{X}_j \cdot \hat{\boldsymbol{\beta}})}}{df} \quad (4.24)$$

The numerator of equation (4.24) represents Pearson's X^2 -statistic. If the assumption of a Poisson distribution of equal mean and variance is satisfied, the expected value of the Pearson X^2 statistic is

equal to the degrees of freedom df . When this is not the case, as in overdispersed Poisson regression, $\hat{\phi}$ is greater than one. This parameter is then used to scale the covariance matrix according to the following equation:

$$Cov_{corr}(\hat{\beta}) = Cov(\hat{\beta}) \cdot \hat{\phi} \quad (4.25)$$

$$se(\hat{\beta}) = \sqrt{Cov_{corr}(\hat{\beta})} \quad (4.26)$$

the se function defined by the right hand side of equation (4.26) yields standards error of the coefficient estimates. After scaling $Cov(\hat{\beta})$, standard errors can be calculated for the regression estimates. In the case of ordinary Poisson regression, the coefficient estimates follow a normal distribution. But since the true variance needs to be estimated from the data, as in ordinary linear regression, the coefficient estimates follow a t-distribution as in ordinary linear regression [121]. Given this fact, upper and lower confidence bounds can be computed as:

$$\hat{\mathbf{LB}} = exp(\hat{\beta} - se(\hat{\beta}) * F_t^{-1}(1 - \frac{\alpha}{2}, df)) \quad (4.27)$$

$$\hat{\mathbf{UB}} = exp(\hat{\beta} + se(\hat{\beta}) * F_t^{-1}(1 - \frac{\alpha}{2}, df)) \quad (4.28)$$

where F_t^{-1} is the inverse cumulative distribution function of the t-distribution. The vector of p-values corresponding to the vector of parameter estimates $\hat{\beta}$ estimate can be correspondingly computed as:

$$\mathbf{p} = 2 * F_t(\frac{|\hat{\beta} - \mathbf{0}|}{se(\hat{\beta})}, df) \quad (4.29)$$

This regression problem of equation (4.20) was implemented in MATLAB, in which the design matrix was manually computed and the final regression problem was solved using *glmfit* function of MATLAB, specifying Poisson regression as input argument. The function has routines to calculate the quantities from above and associated standard errors. For the data set at hand, $\sqrt{\hat{\phi}}$ ranged between 12-16, such that the assumption of an ordinary Poisson process is not satisfied.

For each protein one regression is performed, yielding one p-value and estimate for the FC of this protein between any included condition with respect to the reference. When a total set of various hundred proteins are checked for significantly changed expression, α -error accumulation will lead to a higher number of false-positive differences as the the α -error threshold for each individual protein suggests. Therefore, the set of all obtained p-values was corrected according to [19] to control the α -error rate to 5 % on the entire set of performed significance tests.

11.3.6 Model selection loop to remove non-sensitive factors

Whether a 'replicate' and a 'replicate*peptide' effect is needed to account for differences in absolute fragment ion trace intensities can be answered by a model-selection loop, in which the regression model is successively reduced: First, the interaction effect between peptide and replicate is removed. If this turns out to be non-significant, the main effect is removed and its significance is evaluated. To test for significance of a model term, the residual deviance is used which is a generalization of the residuals sum of squares in ordinary linear regression to the case of Poisson regression. The residual deviance is distributed according to a χ^2 -distribution as is any difference between deviances. Therefore, at each level of the model selection loop the test statistic is computed as:

$$p = 1 - F_{\chi^2}(\text{dev}_{\text{Smaller}} - \text{dev}_{\text{Larger}}, \text{df}_{\text{Smaller}} - \text{df}_{\text{Larger}}) \quad (4.30)$$

where $\text{df}_{\text{Smaller}}$ are the degrees of freedom of the smaller model and F_{χ^2} is the cumulative distribution function of the χ^2 -distribution. When p is smaller than 0.05, all model terms of the larger model are significant and no model reduction can be performed. If not, the next iteration is performed to check if the main effect is needed.

11.3.7 Validation

The above outlined scheme for FC estimation was tested on a series of isotopic standards of known concentration. In particular, Maltose Binding Protein (see subsection 6.3 of the Materials and Methods section) was chosen as model protein, for which ^{14}N - and ^{15}N -standards were commercially available. The ^{14}N - version of this protein was available as 1 mg ml^{-1} solution, which was serially two-fold diluted. From each dilution step of the ^{14}N -standard a sample was prepared by spiking $5 \mu\text{l}$ of the considered dilution into a vial with $1 \mu\text{g}$ trypsin and a $5 \mu\text{l}$ of the ^{15}N -standard of this protein at a concentration of $10 \mu\text{g ml}^{-1}$. The total volume was filled up to $100 \mu\text{l}$ with trypsin digestion buffer. The samples were digested and processed according to Voges and Noack [183] and afterwards injected into the LC-ESI-MS/MS system and measured according to section 7.6 of the Materials and Methods section.

From the set of all extracted mass traces, only those were included into the regression approach which showed a R^2 greater than 0.85 for the $^{14}\text{N}/^{15}\text{N}$ ratio over the entire dilution series because some of the extracted mass traces did not accurately reflect the dilution series. Any threshold stricter than $R^2 > 0.85$ led to reasonable agreement between true and estimated FCs. Figure 4.20 shows the plot of estimated FCs vs true FCs on the linear and the $\log_2$ -scale, respectively. It can be seen that for a certain region the FCs are estimated quite accurately. For lower concentrated samples, however, the FC estimation becomes less accurate and precise. This may be due to the fact that the probability that a fragment ion trace will evaluate to zero increases along the dilution series. Therefore, an ever smaller number of fragment ion trace intensities is available per sample along the dilution series. The non-integrated mass traces are missing data points from the dataset and are set to NaN in the

regression approach. With fewer data points from which the FC can be calculated, the estimation becomes increasingly imprecise.

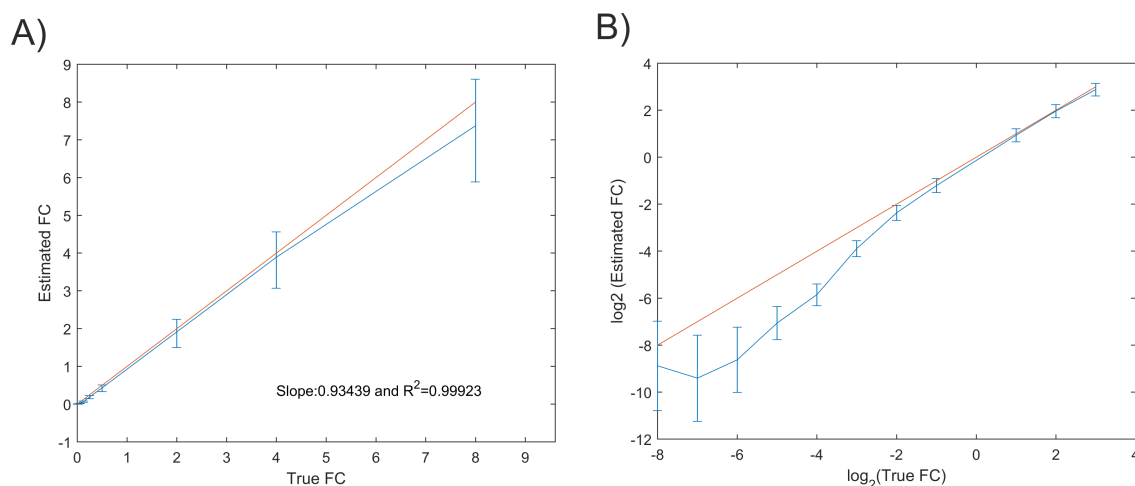


Figure 4.20: Estimated FCs plotted against true FCs as determined by the Poisson regression approach across a dilution series of naturally labeled Maltose binding protein spiked with a constant amount of ¹⁵N-labeled Maltose binding protein. Each data point corresponds to one sample of the dilution series, which was performed serially in two-fold steps. A) Estimated vs. true FCs on the linear scale; B) Estimated vs. true FCs on the log₂-scale. The estimated FCs are indicated together with the confidence bounds calculated according to equations (4.27) and (4.28).

11.4 Fold changes of selected anaplerotic enzymes

As a first quality check on the data set, the estimated FCs of enzymes deleted in some mutants can be considered. It turns out that the corresponding deletions could be verified in each mutant. Exemplarily, the estimated FCs for the enzymes PCx and ME are shown in Figure 4.21. Apparently, the deletion of PCx is identified with high confidence in the respective strains evidenced by the small *p*-values for the corresponding FCs. Except for the strains deficient in PCx, no other significant expression changes of PCx with respect to the WT are apparent across the set of deletion strains. The expression of PCx is known to be subject of regulation by the regulator proteins GlxR, RamB, and SigA [142]. Based on the finding of an unchanged expression, it can be concluded that the observed changes in metabolite pools did not bring about an altered affinity of these transcriptional regulators to the promoter region of the PCx gene. A different picture emerges for ME, the deletion of which is again clearly visible in the corresponding strains. Besides, ME appears down-regulated in the Δ (PCx PEPCx) and Δ PCx strains. Based on this regulation, it appears unlikely that ME serves as anaplerotic enzyme in *C. glutamicum* as shown for the ME of *E. coli* [207]. Since its total specific amount is reduced with respect to the WT and its intracellular substrate pool of PYR is not significantly altered (see section 10), it is unlikely that ME catalyzes a flux compensating for the missing PCx activity in the Δ (PCx PEPCx) and Δ PCx strains. Even with greatly increased intracellular PYR concentration, ME does not appear to be able to catalyze anaplerotic carboxylation for the Δ PCx strain is unable to grow

on PYR as sole carbon source (data not shown). However, Gourdon et al. [63] showed that purified ME of *C. glutamicum* readily carboxylates PYR in *in vitro* assays with an apparent K_m constant of 13.8 mM for PYR. The same study showed that many central carbon metabolites did not affect its activity to a significant extent. Therefore, control of its activity through metabolite effects seems unlikely. More likely, a transcriptional regulatory mechanism appears to repress the expression of the *malE* gene under circumstances when its anaplerotic carboxylation would substitute for missing PCx activity.

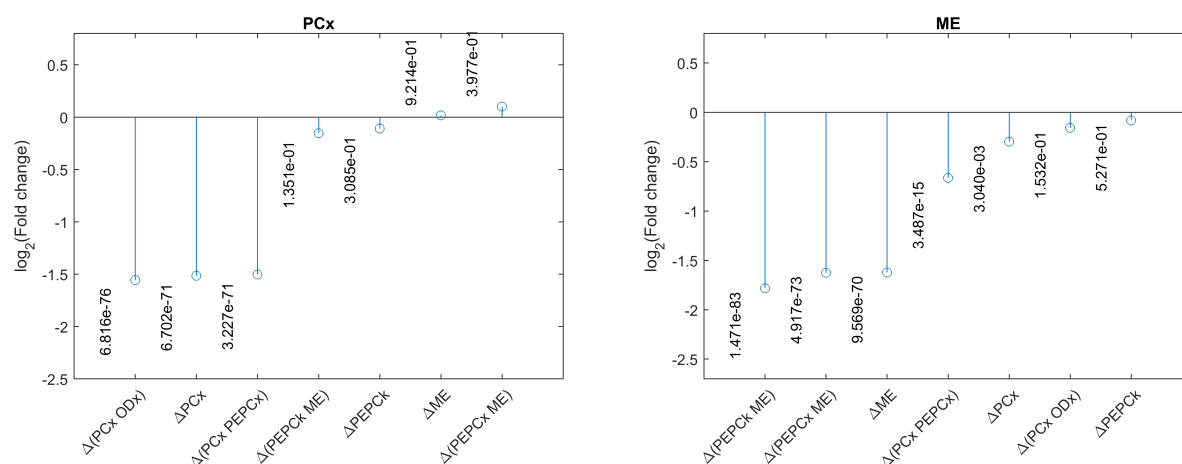


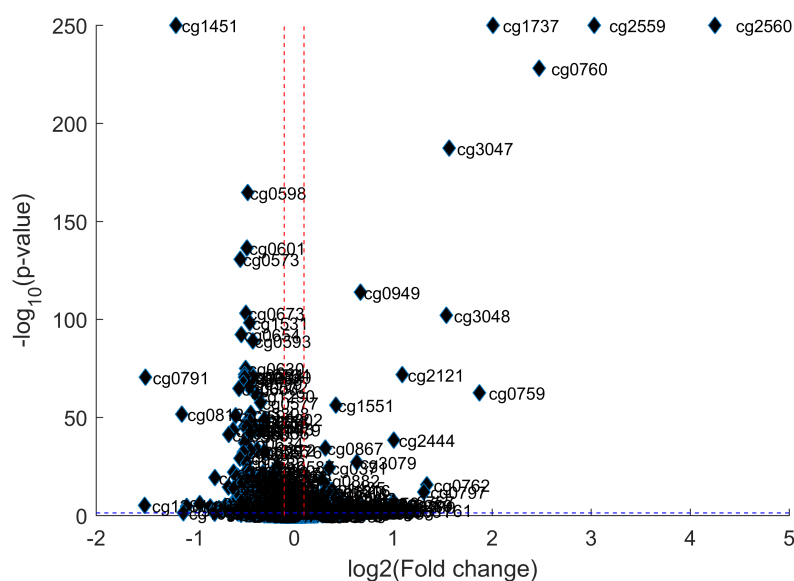
Figure 4.21: FCs of the enzymes PCx and ME across the set of anaplerotic deletion mutants. The height of each stem indicates the magnitude of the FC with respect to the WT. Each stem is annotated with the p -value of the coefficient representing the FC from the regression problem. The shown p -values are corrected for FDR.

11.5 Fold change estimation in anaplerotic deletion mutants

The proteome dataset at hand covers 1122 proteins of *C. glutamicum* for which peak areas are available. Since only proteins with at least two extracted peptides were included into the regression approach, the FCs of 949 proteins with respect to the WT could be computed for all strains. Among the set of deletion mutants, the $\Delta(\text{PCx PEPCx})$ strain exhibits the most differentially abundant proteins, i.e. 289 proteins ($|\log_2(\text{FC})| \geq 0.1$, $p < 0.05$). Since the FDR was controlled at 5 % over the total set of proteins according to the procedure of Benjamini and Hochberg [19], the total set of significantly altered genes can be expected to contain 5 % false positive identifications. Without FDR control this rate would have been much higher due to α -error inflation. The so determined total set of differentially expressed proteins in the $\Delta(\text{PCx PEPCx})$ strain is shown in Figure 4.22 as volcano plot.

From Figure 4.22 it becomes apparent that the p -value of the FC of ICL (cg2560) in the $\Delta(\text{PCx PEPCx})$ mutant clearly suggests an up-regulation. The same holds true for the second enzyme of the GLX shunt, i.e. MS (cg2559). The corresponding genes of both enzymes, *aceA* and *aceB*⁶ are thought

⁶Both genes are transcribed as monocistronic transcript [146].



to be repressed by the regulator proteins GlxR (cg0350) [141] and RamB (cg0444, *ramB*) [12] under glycolytic conditions. Whereas GlxR and RamB repress the glyoxylate bypass, it has been established that RamA (*ramA*, cg2831) acts as transcriptional activator of both genes when acetate is the carbon source. Apparently, the balance of transcriptional repression and activation is altered in the Δ (PCx PEPCx) mutant in such a way as to effect an increased expression of both genes. The whole genome sequencing results summarized in Table 4.5 did not reveal any SNP in the promoter regions or genes of either *ramB* or *ramA*. Therefore, in the absence of post-translational modifications their function as repressor and activator, respectively, and affinity to possible metabolite effectors can be assumed to be unaffected. Just as little evidence for a mutation in the promoter regions of *aceA* and *aceB* or the genes themselves was found. This makes the conclusion compelling that an altered concentration of a metabolite effector must have shifted the balance of activation and repression of *aceA* and *aceB* through binding to either RamA or RamB.

shown) which is also transcriptionally activated by RamA but could not be shown to be affected by RamB [12]. Despite RamA being active as transcriptional regulator at the the aforementioned genes, not all of its regulon shows a behavior that is concordant with its known role as transcriptional regulator. According to the above results the expression of ME appears to be repressed in the $\Delta(\text{PCx PEPCx})$ mutant, although RamA is thought to be a transcriptional activator of the *malE* gene [12]. However, the *malE* gene is additionally negatively regulated by the regulator MalR [104]. Apparently, this regulator appears to be activated in its repressive function when RamA is active as transcriptional activator for the expression of ME is most strongly suppressed in the mutant with highest apparent RamA activity, i.e. $\Delta(\text{PCx PEPCx})$. This assumption is substantiated by the evidence that ME specific activity in the WT strain is greatly reduced when acetate is the carbon source in comparison to glucose metabolism [63, 104].

The question arises which metabolite effector is altered in concentration in this mutant to cause the binding of RamA to the promoter regions of *aceA*, *aceB* and the *ack-pta* operon to change. To identify the metabolite effector mostly likely to interact with RamA to bring about the transcriptional activation of the *aceA* and *aceB* genes, the set of significantly accumulated metabolite pools for the $\Delta(\text{PCx PEPCx})$ mutant can be considered. Because this set contains several upregulated metabolites pools, no conclusion can be drawn with respect to this question using this point of view. A more fruitful approach consists in considering the FCs of these enzymes across the entire set of mutants. As it turns out the enzymes ICL, MS, ACK, PTA are differentially expressed across the set of analyzed deletion mutants and can be used to group the mutants into two clusters. From Figure 4.23 it appears that the set of anaplerotic deletion mutants can be grouped according to whether the aforementioned enzymes are up- or down-regulated. For cg0949 (CS) the same grouping based on up- and down-regulation was observed. To identify the metabolite effector that regulates the binding of RamA to the corresponding genes, the metabolite pool whose FCs across the set of strains most closely correlates with the ones of Figure 4.23 can be considered. However, from the set of measured metabolites (Figures 4.13 and 4.14), no metabolite pool becomes apparent that exhibits the same ranking of strains based the FCs with respect to the WT. Surely, that does not invalidate the hypothesis of a changed metabolite effector. It may simply be a metabolite pool not included into the targeted metabolomics method. This illustrates the need to extend metabolome acquisition methods to match the almost genome-wide coverage of SWATH-MS proteomics methods.

Of note, the shown FCs of ICL and MS in Figure 4.23 show the same ranking of mutants. The same holds true for ACK and PTA. Moreover, the estimated magnitude of FCs of the latter pair are correlated to each other with an R^2 of 0.996, the ones of ICL and MS with $R^2=0.999$. This is not surprising since ACK and PTA are transcribed as one operon [146]. In the absence of post-translational modifications, the FCs of the translated proteins need to be correlated. This fact also shows the power of the applied approach, which could successfully reproduce this known pattern of transcriptional regulation.

The up-regulated genes of ICL, MS, ACK and PTA in the $\Delta(\text{PCx PEPCx})$ mutant are actually described as hallmark genes of acetate metabolism in *C. glutamicum* WT, under which condition they

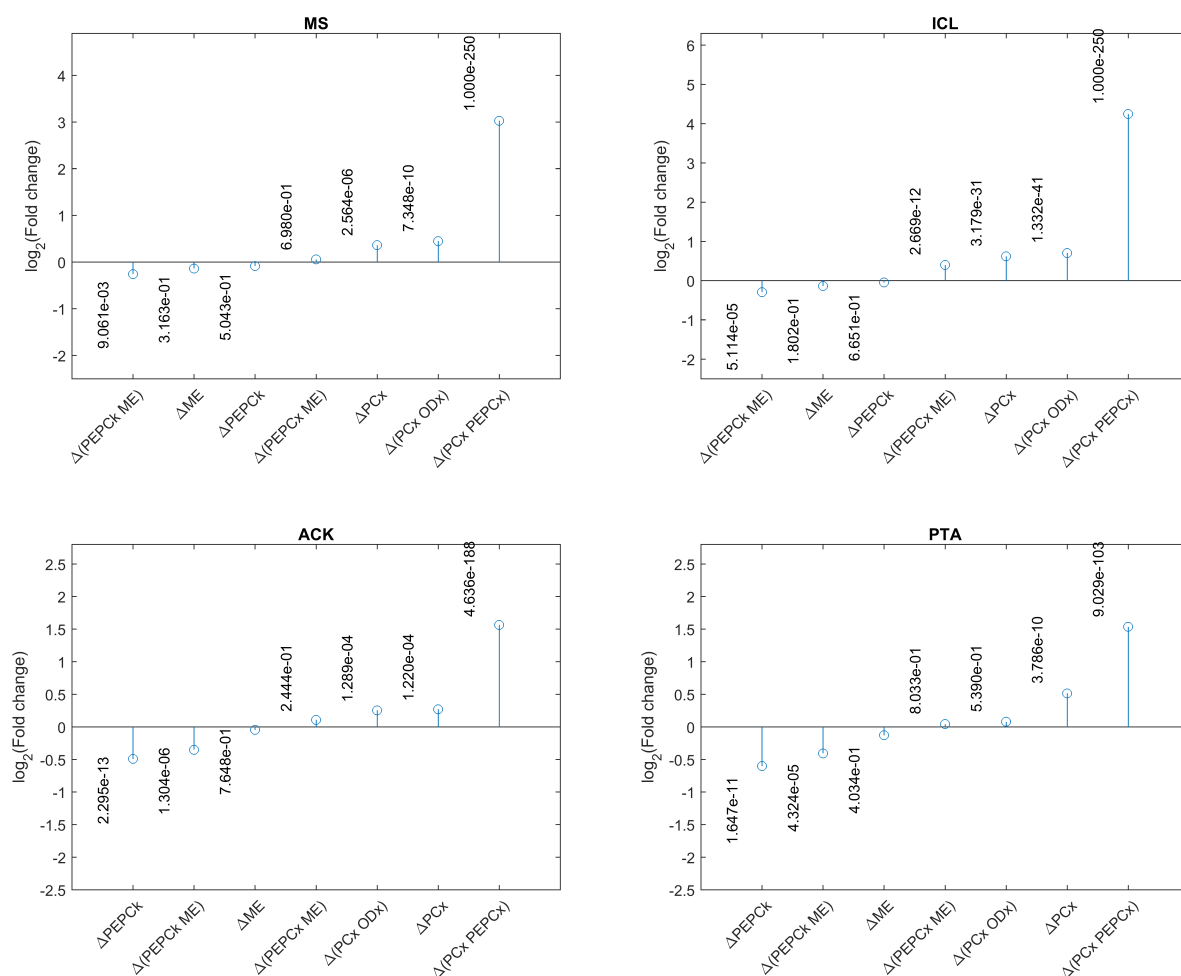


Figure 4.23: The FCs for the products of the genes cg2559, cg2560, cg3047 and cg3048 corresponding to MS, ICL, ACK and PTA. The height of each stem indicates the magnitude of the FC with respect to the WT. Each stem is annotated with the p -value of the coefficient representing the FC from the regression problem. The shown p -values are corrected for FDR.

are markedly up-regulated [172]. For all the apparent similarities of the expression adaptation of the Δ (PCx PEPCx) mutant to the expression profile of *C. glutamicum* WT during growth on acetate, a major difference remains: Cramer and Eikmanns [35] reported that RamA is about three times more abundant in *C. glutamicum* WT grown on acetate as sole carbon source in comparison to glycolytic metabolism, whereas RamB was shown to be two-fold down-regulated under the same condition [36]. Although both studies reported semi-quantitative results based on Western-blotting, the changes are clearly visible in the reported Western blots. Thus, both regulators show differential expression with respect to the carbon source. The proteome data set at hand does not show any significant expression change of either RamA or RamB in comparison to the WT. For RamA this result is based on 51 fragment ion trace intensities from five peptides, whereas the relative quantification of RamB draws on 35 fragment ion traces from three peptides. Therefore, this result can be considered valid. Apparently, the increased expression of RamA under acetate metabolism must be brought about by another regu-

lator that is specific to acetate as carbon source and which is not involved in the $\Delta(\text{PCx PEPCx})$ strain under glucose metabolism. This conclusion is concordant with Cramer and Eikmanns [35] who concluded that the increased expression of RamA under acetate must be due to another regulator and is independent of RamA itself. This view is clearly supported by the presented analysis since the carbon source was not changed with respect to the WT and neither RamA nor RamB show altered expression in comparison to the WT although their target genes do.

As shown above some genes of the known regulon of RamA (ICL, MS, PTA and AK) are up-regulated in the $\Delta(\text{PCx PEPCx})$ mutant. These genes represent just a small subset of the RamA regulon. On the other hand, RamA is just one of many regulators of *C. glutamicum*. The downloadable database CoryneRegNet, summarizing the known transcriptional regulatory interactions, contains 75 distinct regulator proteins for *C. glutamicum* [142]. For each regulator protein, Figure 4.24 depicts the share of significantly altered proteins pertaining to its regulon relative to the total size of its regulon for each anaplerotic deletion mutant. A protein is annotated with its cg-number when more than 20 % of its regulon are significantly changed in at least one anaplerotic deletion strain. It becomes clear that in no mutant all proteins of the regulon of the considered regulator proteins are significantly altered. Even for RamA only a small share of genes is significantly altered in the $\Delta(\text{PCx PEPCx})$ strains. This suggests that the affinity of binding of RamA may be regulated by various metabolites and not a single one.

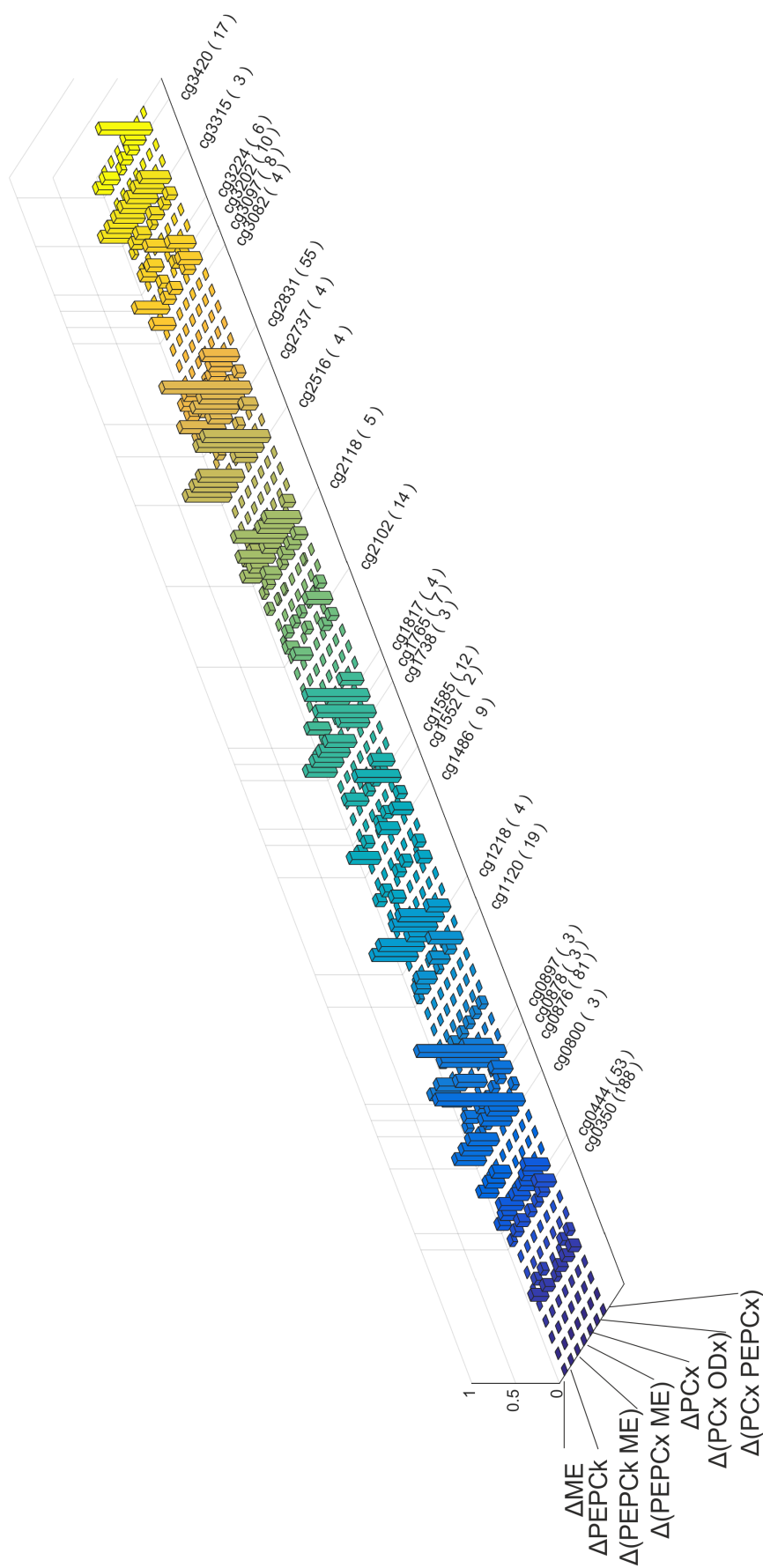


Figure 4.24: Share of proteins of the known regulons of *C. glutamicum* that show significantly altered expression with respect to the WT in relation to the total number of proteins of the respective regulon. Each known regulator is annotated by its cg-number when at least 20 % of its regulon is significantly changed in at least one mutant strain. The cg-number is followed by the total number of regulated loci of the respective regulator in parenthesis. cg0350: glxR, cg0444: ramB, cg0800: prpR, cg0876: sigH, cg0878: whcE, cg0897: pdxR, cg120: ripA, cg1218: nrtR, cg1486: ltbR, cg1552: qorR, cg1585: argR, cg1738: acnR, cg1765: sufR, cg1817: pyrR, cg2102: sigB, cg2118: fruR, cg2516: hrcA, cg2737: fasR, cg2783: gntR1, cg2831: ramA, cg3097: hspR, cg3202: farR, cg3224: lldR, cg3315: malR, cg3420: sigM

12 ¹³C-Metabolic Flux Analysis

12.1 ¹³C-Labeling experiment and enrichment profiling

The set of anaplerotic deletion mutants was further investigated by performing a dedicated ¹³C-labeling experiment. The substrate isotopomer mixture for this experiment consisted of 3.35 mol% naturally labeled glucose, 81.2 mol% [1-¹³C]glucose and 15.4 mol% [U-¹³C]glucose. This mixture was determined by averaging the optimal substrate isotopomer mixtures for individual anaplerotic mutants determined by an *a priori* experimental design using the *edscanner* function of 13CFLUX2. It may be noted that the strain-wise optimal mixtures did not differ to a significant extent. The experiment was performed equivalently to the metabolome experiment described in section 10, except that the carbon source consisted of the above mentioned isotopomer mixture at the same molar concentration as the glucose concentration employed in the metabolome experiment. The ¹³C-enrichment was determined using the acquisition methods described in the Material and Methods section of this work (subsection 7.5).

The resulting fractional enrichment of the mass traces of the PEP and ASP pools is shown in Figure 4.25A. Apparently, only slight differences in the mean fractional enrichment can be observed for metabolites whose labeling pattern ought to change most significantly as they are product pools of reactions deleted in the strains. To judge if the labeling pattern of the above mentioned metabolites actually changed with respect to the strain background in a statistically significant way, a multivariate 8-way ANOVA appeared appropriate since the labeling pattern is represented by at least $\kappa^C + 1$ measured mass traces, of which κ^C are independent, when κ^C denotes the number of carbon atoms of the metabolite in question. In this analysis, a synthetic new variable is searched for as linear combination of the observed variables, along which the F-statistic of an univariate ANOVA (using the known grouping of the samples) is maximized. Analogously, a second new variable can be constructed which is orthogonal to the first one. Each sample can then be plotted in the space of the transformed variables, which are denoted first and second canonical variables (Figure 4.25 B). It appears that although the strain differences on each individual mass trace are barely distinguishable, the overall differences in the labeling pattern, captured by the multivariate ANOVA, are still highly significant. It turns out that the labeling pattern is significantly different for nearly all measured metabolites except GLY, SER, TRP and E4P.

It can be expected that the removal of a reaction at the anaplerotic node should reflect itself most clearly on the labeling pattern of the product of its catalyzed reaction since one of its label influxes is removed. Clearly, this is only the case when the removed reaction catalyzes an *in vivo* flux large enough to lead to a distinguishable contribution to the labeling pattern. By this token, the apparent clustering on the first canonical variable of ASP can be explained (Figure 4.25 B). ASP derives its labeling linearly from OAA, which was experimentally not accessible, and therefore its labeling pattern can be taken as representative for that of OAA. OAA is the product of the reaction catalyzed by PCx, missing in the $\Delta(\text{PCx ODx})$ and ΔPCx mutants. Conspicuously, the samples of those two mutants

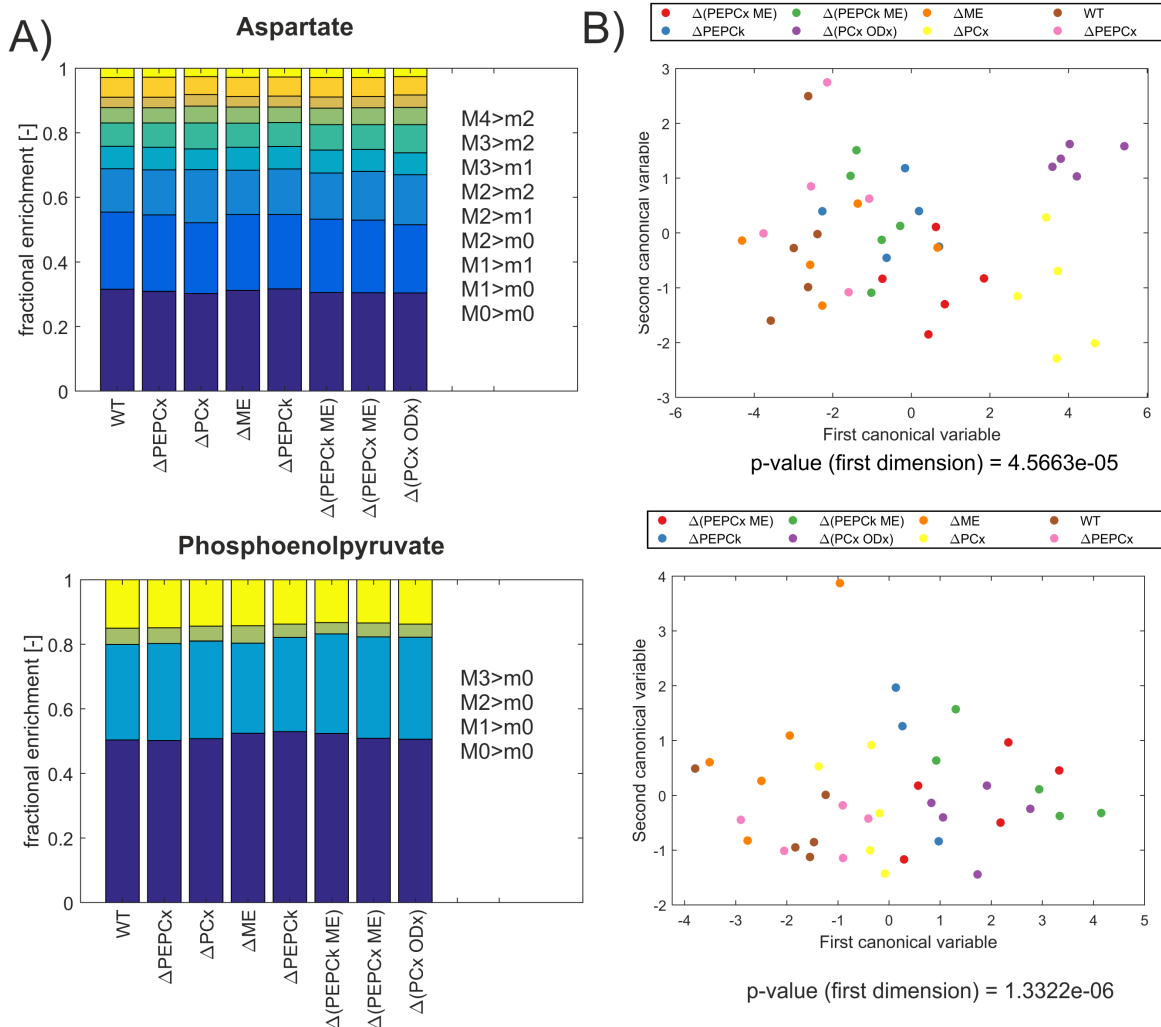


Figure 4.25: ^{13}C -label enrichment in selected metabolite pools and statistical analysis of the labeling pattern. A) The magnitude of each stack of one bar represents the magnitude of the relative mass trace intensity. The fractional abundances of individual mass traces are ordered according to the ordering of mass trace denominators on the right to each bar plot. The plotted quantity represents the mean from averaging the fractional mass trace abundance over five quenching replicates. For some strains, fewer replicates were available per metabolite when the metabolite in question eluted outside of the elution time window in the scheduled MRM assay. From bottom to top, the fractional enrichment of each mass trace is plotted. The mass traces are ordered by the number of labeled carbon atoms. B) Each sample is plotted in the space of the first and second canonical variable from the linear discriminant analysis (multivariate ANOVA) on the first κ^C measured mass traces.

cluster together on the first canonical variables.

On the other hand, the reaction of PEPCx also catalyzes a reaction of which OAA (represented by ASP) is the product. The two deletion mutants comprising these deletion, i.e. Δ PEPCx and Δ (PEPCx ME), do not separate from the WT samples in the space of the canonical variables. Since a reaction affects a labeling pattern only if it is indeed active or significant, PEPCx appears to catalyze a far smaller flux in comparison to PCx. This view is fully consistent with Peters-Wendisch et al. [143] who identified that PCx catalyzes the gross of anaplerotic carboxylation and all previous data presented in this thesis. This is supported by the statistical analysis of the labeling pattern of ASP.

The mutant strains Δ PEPCk and Δ (PEPCk ME) are missing the reaction from OAA to PEP. Since this reaction is indeed active under *in vivo* conditions in the WT (see section 9), the labeling pattern in PEP should discriminate those two strains from all other. As apparent from the multivariate ANOVA result in Figure 4.25 both strains cluster together on the positive axis of the first canonical variable. However, both cluster together with the Δ (PCx ODx) and Δ (PEPCx ME) strains on the positive axis. This shows that their labeling pattern is more similar to the mutants deficient in PEPCk than to other strains and raises the expectation that PEPCk activity is lower in the Δ (PCx ODx) and Δ (PEPCx ME) mutants compared to other strains with which they do not cluster.

The labeling pattern in PYR is depicted in Figure 4.26. It appears that all deletion mutants involving a gene deletion in ME, i.e. Δ ME, Δ (PEPCk ME) and Δ (PEPCx ME), cluster on the negative axis of the first canonical variable. Therefore, by the arguments from above, ME must catalyze a distinguishable flux under *in vivo* conditions towards PYR in the WT, since the labeling pattern of the product should change most clearly to the removal of a reaction that feeds into it. However, it appears that all those mutants also cluster in the lower right corner in the plot of principle components of MAL. By the argument from above, this would mean that also the reverse flux must be prevalent under *in vivo* conditions.

12.2 Flux estimation

A flux estimation based on the obtained ¹³C-labeling data was performed using the network of section 8.2. Each optimization run was initialized at a uniformly-random sampled point from the flux polyhedron using the *ssampler* function of the 13CFLUX2 software package. For each network 500 optimization runs were initialized in this way, which were carried out using the NAG NLP solver from the 13CFLUX2 package, employing the default criteria for optimization termination. The final solutions were transformed from the net and exchange-flux coordinates used in 13CFLUX2 into forward and backward flux coordinates according to Wiechert and de Graaf [193]. The frequency distribution of the final optimizer output for the 500 starting points is shown in Figures 4.27 and 4.28 for anaplerotic reactions as well as two fitted specific extracellular rates, the specific glucose uptake rate and the biomass efflux from ACCOA. Contrary to the network described in section 8.2, the biomass efflux reaction from PEP and ACCOA were not constrained to the value calculated from the employed biomass composition. This proved to be necessary since the distributions of final optimizer outputs

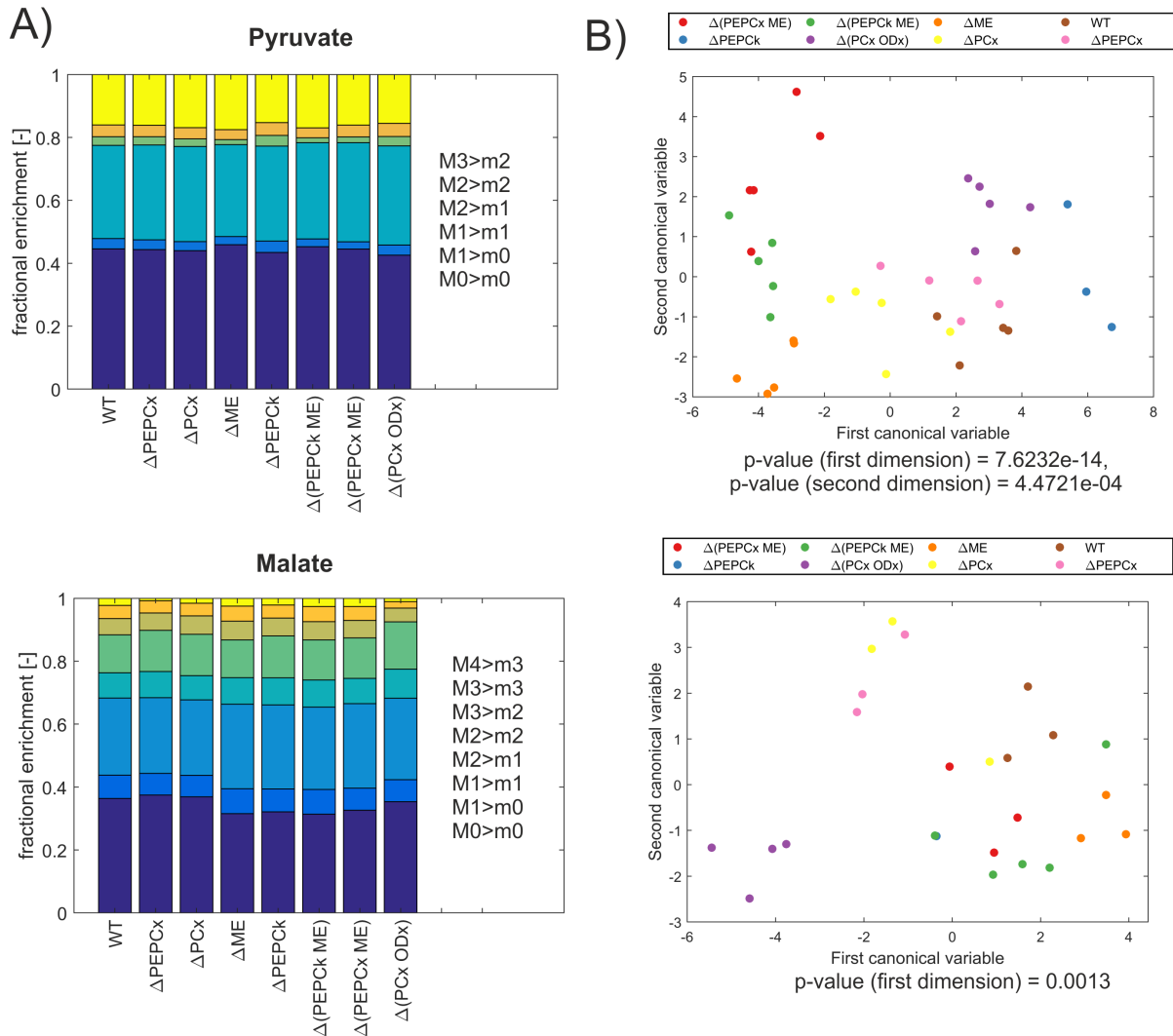


Figure 4.26: ^{13}C -label enrichment in metabolite pools A) PYR and B) MAL. From bottom to top the fractional enrichment of each mass trace is plotted, ordered by the number of labeled carbon atoms. For the Δ PEPCk strain only one measurement of MAL was available. The corresponding sample has almost identical coordinates as the leftmost sample from the Δ (PEPCk ME) which partly covers the corresponding dot.

were bimodal for some fluxes otherwise (data not shown).

Expectedly, no unique solution could be found for some anaplerotic fluxes of the WT network. Notwithstanding, an unique solution was obtained for the PEPCk flux in the WT, which is consistent with the conclusions from section 9.3. However, the solutions for selected fluxes of other mutants do not meet the benchmarks derived from the extracellular specific rate measurements, one of which stipulates that ME must be active in the Δ PEPCk mutant. Contrary to this, the vast majority of solutions for both forward and backward flux of ME converged towards 0 in the Δ PEPCk mutant. Furthermore, the obtained solutions for the Δ PCx mutant yield a flux value for ME that is greater than that of PEPCx in this mutant and greater as the ME flux in other strains, suggesting that ME compensates for the missing anaplerotic activity of PCx in this strain. However, this contradicts the suggested

flux regulation inferred from the proteome data from section 11.4 and the fact that extracellular rate measurements suggest an important role for PEPCx in this mutant.

Moreover, the estimated specific glucose uptake rate is not consistent with the extracellular specific rates from section 9.3. The fitted specific uptake rates greatly overestimate the measured uptake rates in the ΔPEPCx , ΔPCx , $\Delta(\text{PEPCx ME})$ strains. Furthermore, the histogram of the estimated biomass effluxes from the ACCOA pool has a bimodal distribution for the WT strain. The mode at $1.5 \text{ mmol (g CDW h)}^{-1}$ approximately matches the value this flux would have been constrained to, if the full biomass equation had been employed, i.e. $1.946 \text{ mmol (g CDW h)}^{-1}$ for a growth rate of 0.45 h^{-1} . However, for all other mutants the biomass flux from ACCOA converges towards zero, which is clearly not consistent with the underlying metabolism. The biomass efflux from ACCOA mainly feeds into the biosynthesis of mycolic acids which constitute an essential building block of the *C. glutamicum* cell wall. Since no growth defects were detected for several mutant strains, it appears very unlikely that this flux is an order of magnitude lower in comparison to the WT in those mutant strains.

Most importantly, an identical measurement setup was used for the flux inference in each strain. Therefore, the sum of squared residuals (SSR) of the solutions of different strains should be identical up to experimental error, if each network topology indeed represents the metabolism that generated the measured label distribution. Figure 2 in Appendix 8 shows the histogram of the residuals of the solutions for each strain. Apparently, the SSR varies drastically between the investigated strains, most significantly so in the ΔPEPCk mutant. Therefore, it can be inferred that the change of metabolism introduced through the deletion is not properly accounted for by solely removing the carbon atom transitions of the corresponding reactions from the network of section 8.2.

The space of possible error sources explaining the above finding is too broad to explore in every aspect. Flawed measurements of ^{13}C -enrichment are certainly to blame. In fact, the following section will show that the ^{13}C -enrichment profiling in sugar phosphates according to the method of section 7.5 is biased by approximately 5 mol% (Table 5.1). Most likely, this effect is due to an unspecific retention and elution of this metabolite class on the used column material, possibly mediated by the employed ion pairing reagent. Although this effect will contribute to a biased flux map explaining the discrepancies to the conclusions from the orthogonal measurement of extracellular rate measurements, it cannot account for the discrepancies across the set of investigated strains. The biased ^{13}C -enrichment measurement will affect the labeling pattern of all deletion strains to the same extent. Rather, the assumption of an equal composition of biomass dry matter in all deletion strains may be violated. This assumption is customary in ^{13}C -MFA, not least due to the difficulty of biomass composition analysis. Recently, Long and Antoniewicz [112] investigated the variability of biomass dry matter composition in *E. coli* with respect to changes in the genetic background. The authors could show that a deletion of the Phosphoglucose isomerase reaction of *E. coli* led to a change of CDW content of almost 10% for selected constituents of cell dry matter. This suggests that an altered biomass composition may further contribute to the observed inconsistencies. The consequences of an altered biomass composition for anaplerotic reactions can be illustrated with equation (4.31), which arises

from the stoichiometry of the WT network of section 8.2:

$$\begin{aligned}
 v_{\text{PEPCx_PEPCK}}^{\text{net}} + v_{\text{ME}}^{\text{net}} + v_{\text{PCx_ODx}}^{\text{net}} = & v_{\text{bmARG}}^{\text{net}} + v_{\text{bmASN}}^{\text{net}} + v_{\text{bmASP}}^{\text{net}} \\
 & + v_{\text{bmDAP}}^{\text{net}} + v_{\text{bmGLN}}^{\text{net}} + v_{\text{bmGLU}}^{\text{net}} \\
 & + v_{\text{bmILE}}^{\text{net}} + v_{\text{bmLYS}}^{\text{net}} + v_{\text{bmMET}}^{\text{net}} \\
 & + v_{\text{bmPRO}}^{\text{net}} + v_{\text{bmTHR}}^{\text{net}} + v_{\text{bmUMP}}^{\text{net}}
 \end{aligned} \tag{4.31}$$

Analogous relations can be derived for other mutants, which all constrain the sum of anaplerotic net fluxes to a fixed value, dependent only on the growth rate. If the specific amounts of these amino acids are different across the set of strains, however, a wrong constraint is introduced on the anaplerotic net activity, forcing the corresponding fluxes to a value discordant with the label measurements.

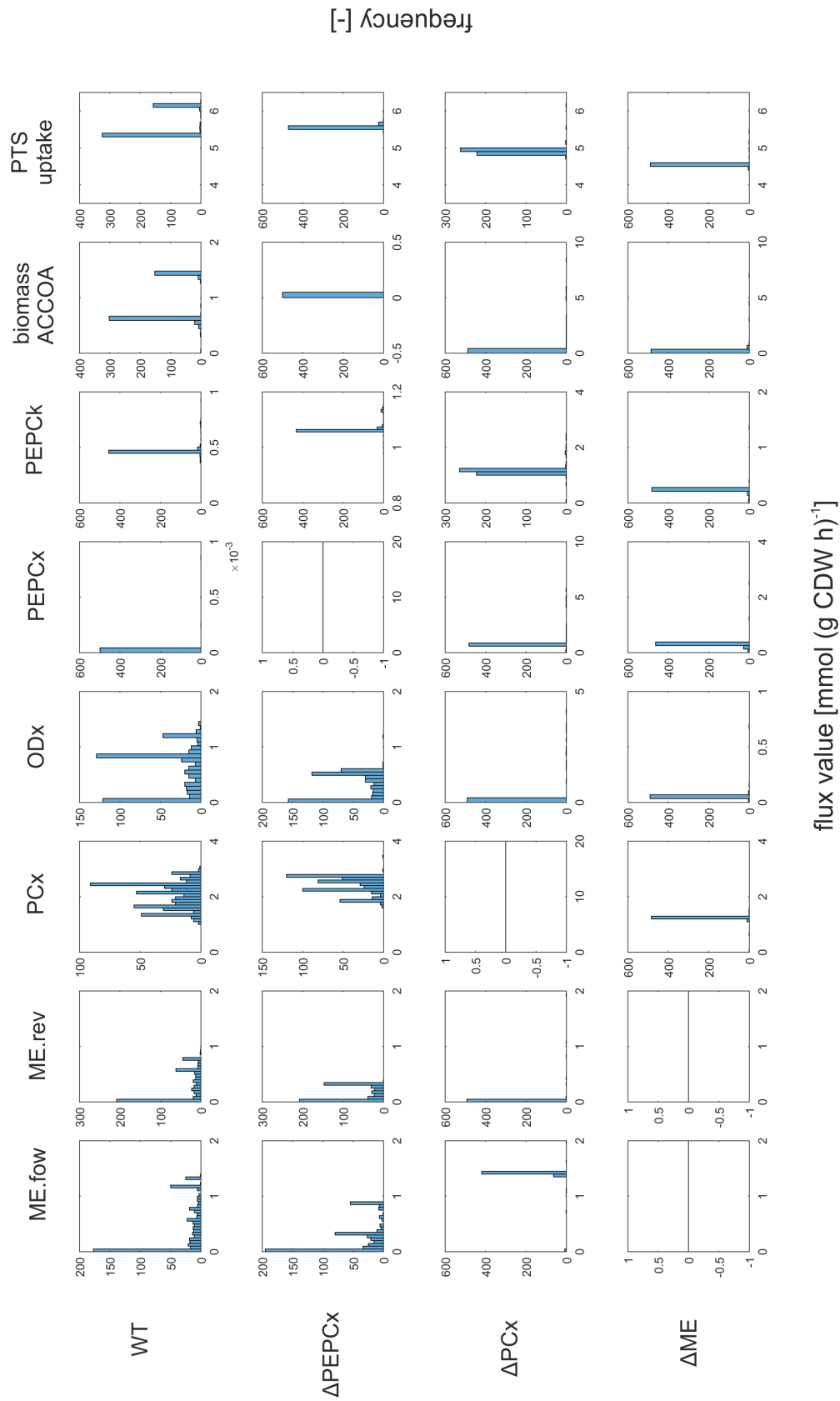


Figure 4.27: Estimated flux values for WT, Δ PEPCx, Δ PCx and Δ ME strains. Each histogram depicts the frequency of individual solutions for the indicated fluxes from 500 optimization runs which were randomly initialized.

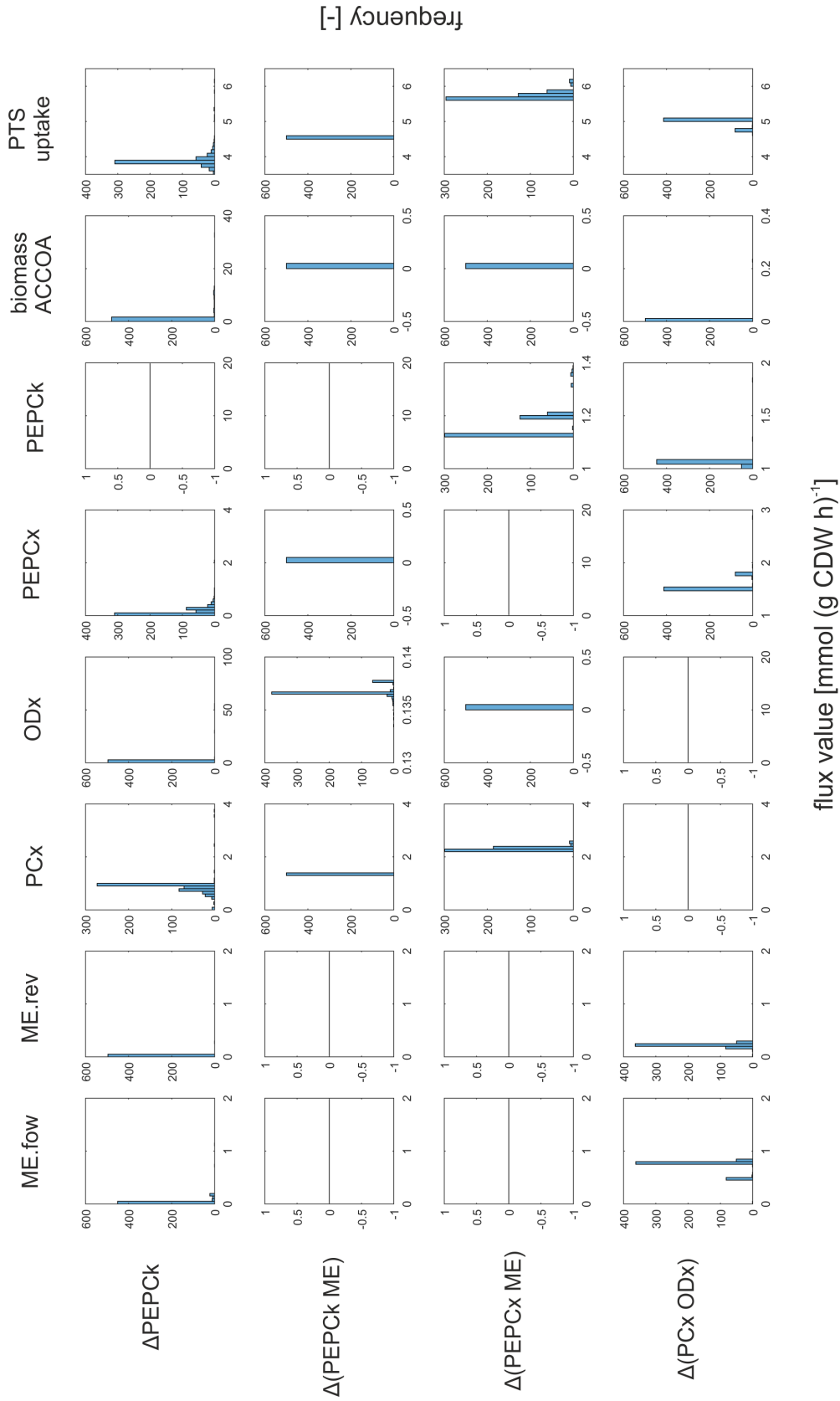


Figure 4.28: Estimated flux values for Δ PEPCK, Δ (PEPCK ME), Δ (PEPCx ME) and Δ (PCx ODx) strains. Each histogram depicts the frequency of individual solutions for the indicated fluxes from 500 optimization runs which were randomly initialized.

Results and Discussion - Part 2

13 QqTOF MS technology for ^{13}C -isotopomer profiling

13.1 Motivation

The previous section highlighted some limitations inherent in the established metabolomics and ^{13}C -enrichment profiling methods based on QqQ MS. In particular, this platform inherently limits the number of tandem mass isotopomer transitions that can be simultaneously monitored in the time scale of a HPLC elution. On the other hand, novel QqTOF MS devices acquire quantitative, accurate product ion spectra at high acquisition rates without cycle time limitation. These features and the ambiguous results from the preceding section motivated the development of a QqTOF MS-based ^{13}C -profiling method, harvesting the unique properties of QqTOF MS technology. First, an acquisition method is developed to separate samples using LC and acquire product ion spectra thereof. A customizable data processing pipeline is developed which allows for efficient data extraction. This scheme is applied to an experimental acquisition of an isotope dilution standard series. However, to harvest the positional information contained in the product ion spectra for improved isotopomer resolution for ^{13}C -MFA, the positional identity of product ion carbon atoms needs to be known. Therefore, the second part of this chapter is concerned with the structural elucidation of product ion structures. Finally, with the capabilities of QqTOF MS technology for ^{13}C -isotope profiling demonstrated, the use of this technology for analytical separation of ^{13}C - and ^{15}N -MI species will be explored.

13.2 ^{13}C -profiling method

13.2.1 Liquid chromatography and acquisition method

Since ion pairing reagents are known to mediate memory effects in liquid chromatography, in particular visible in the data of Table 5.1 of section 13.3.4, a novel HPLC method was implemented. As an alternative to ion pairing reagents, hydrophilic interaction chromatography has been used in several studies for separation of the primary metabolome [205, 13, 170]. For hydrophilic interaction chromatography a multitude of functionalized stationary phases are available, such as those bearing amino, amide or zwitterionic groups. The polymer-based column material functionalized with a sulfobetaine zwitterionic group used in the study of Teleki et al. [170] for separation of primary metabolites was shown to resolve many isobaric metabolites such as CIT and ICIT as well as ISOLEU and LEU under alkaline conditions. It was therefore chosen as basis for a LC method which employs 10 mM ammonium acetate (pH 9.2) as eluent A and acetonitrile as eluent B. For metabolite quantification using a fully-labeled IS, the gradient eluent from Teleki et al. [170] can be used. When ^{13}C -enrichment is profiled the eluent gradient of Table 7 in Appendix 9 is used to account for the increased cycle time resulting from the large number of Q1 windows that have to be monitored simultaneously.

During the elution, the Q1 windows corresponding to carbon-labeling states of included precur-

sor ions are monitored using an optimal DP value. To maximize sensitivity, the CE-parameter is usually set to the value optimizing the signal intensity of the most intense product ion, identified from a CE-ramp. On the other hand, the maximal fragmentation is achieved by ramping through a large domain of CE values. As a compromise between extensive fragmentation and sensitivity, a CE-ramp of ± 15 eV is performed in each Q1 window centered on the CE value of the most intense fragment.

During acquisition the mass axis of the TOF detector was calibrated every second sample using the calibration solutions reported in section 6.4. The calibration was carried out using the auto-calibration routine of the device. For optimal accurate mass stability during acquisition, the choice of precursor ion from the calibration solution for MS/MS calibration proved crucial and is therefore reported at this point. In positive ionization mode the m/z 315.1623 precursor ion was used for MS/MS calibration, yielding fragment masses from m/z 315.1623 down to m/z 58.0651. For negative ionization mode calibration the m/z 277.0983 precursor ion was chosen which generated fragment masses from m/z 277.0983 down to m/z 77.0397. This choice of calibration ions resulted in fragment ions that allowed an accurate calibration of the mass range relevant for metabolomics without derivatization ($m/z=30-180$).

13.2.2 Automatic computation of mass traces

Ordinarily, acquisition methods for ^{13}C -isotopomer profiling acquire the TMID from one or limited number of fragments. In this case, the number of mass transitions to be monitored or extracted from the product ion spectra is limited and can be specified manually. If an increased positional resolution of ^{13}C -label is desired, the TMIDs of an increasing number of fragment ions need to be acquired, translating itself into an equally higher number of mass transitions which makes the manual specification of fragment ion traces to be extracted infeasible. Therefore, a method was implemented that automatically computes the arising ^{13}C -labeling states for a predefined set of product ions per metabolite, which is used to automatically generate a quantitation method as text-file that can be imported to the manufacturer software MultiQuant for XIC integration.

To formalize this computation, the notation outlined in Figure 5.1 will be used. In this figure the variable $\kappa^{\text{P},\text{C}}$ refers to the number of carbon atoms of a given precursor ion, which is identical to the number of carbon atoms of the corresponding metabolite. Analogously, the variable $\kappa^{\text{F},\text{C}}$ denotes the number of carbon atoms in the fragment ion. The Greek letter λ will denote the labeling state of a chemical structure. Thus, to denote the carbon labeling state of a pair of precursor and fragment ions, the variables $\lambda^{\text{P},\text{C}}$ and $\lambda^{\text{F},\text{C}}$, respectively, are used. Consequently, the notation of a mass trace corresponding to a carbon labeling state can be generalized to $M[\lambda^{\text{P},\text{C}}] > m[\lambda^{\text{F},\text{C}}]$.

To compute the exact mass transitions corresponding to ^{13}C -labeling states of each precursor ion MI with $\lambda^{\text{P},\text{C}}$ labeled carbon atoms, the set of possible labeling states in a chosen fragment ion with $\kappa^{\text{F},\text{C}}$ carbon atoms must be computed. This set will be henceforth denoted $TMI_C(\lambda^{\text{P},\text{C}}, \kappa^{\text{P},\text{C}}, \kappa^{\text{F},\text{C}})$. For fragments originating from a carbon backbone break, the number of eliminated carbon atoms during fragmentation is equal to $\kappa^{\text{P},\text{C}} - \kappa^{\text{F},\text{C}}$. Depending on whether the ^{13}C -label is present in the

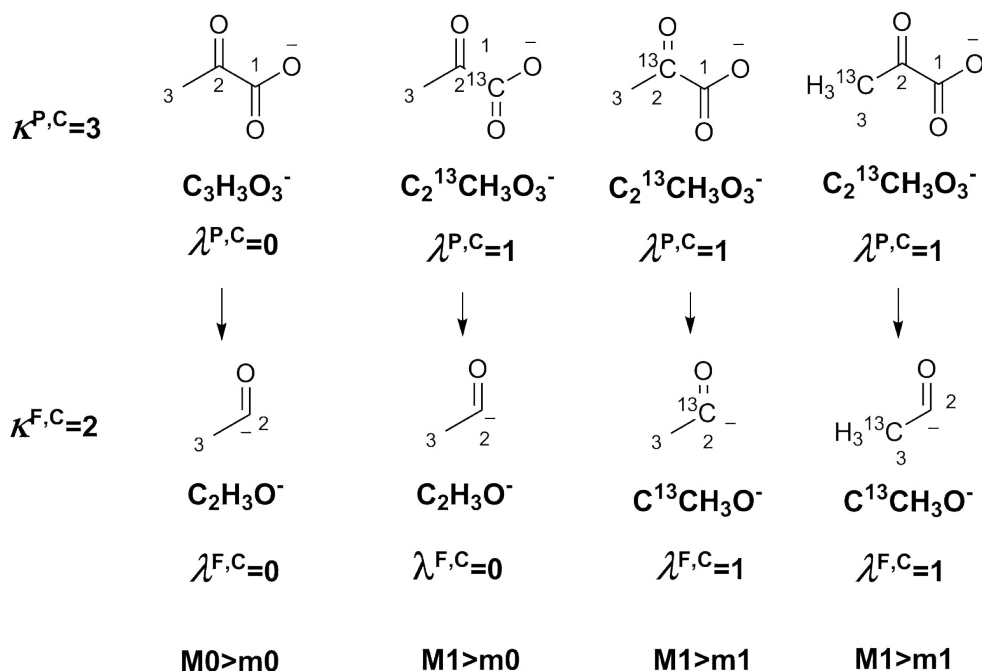


Figure 5.1: Notation for carbon labeling states of precursor and fragment ions used in this thesis. The isotopomers of PYR which contribute to the mass traces M0>m0, M1>m0 and M1>m1 of fragment m/z 43.0189 are shown. On the left, the variables which denote the number of carbon atoms are shown: $\kappa^{P,C}$ refers to the number of carbon atoms in the precursor ion, whereas the variable $\kappa^{F,C}$ denotes the number of carbon atoms in a fragment ion. Below each precursor ion MI the variable $\lambda^{P,C}$ denotes the number of labeled carbon atoms in the precursor ion, i.e. the mass isotopomer of the metabolite in question. $\lambda^{F,C}$ denotes a tandem mass isotopomer, i.e. the number of labeled fragment ion carbon atoms.

eliminated carbon atoms, the fragment ion labeling states $\lambda^{F,C}$ may be smaller than that of the precursor $\lambda^{P,C}$. Therefore, the minimal labeling state in product ions with fewer carbon atoms than the precursor is $\lambda^{P,C} - (\kappa^{P,C} - \kappa^{F,C})$. When the number of labeled carbon atoms in the precursor $\lambda^{P,C}$ is smaller than number of lost carbon atoms during CID, the lowest possible labeling state of the fragment is 0. Its maximal labeling state, on the other hand, is equal to the one of the precursor ion $\lambda^{P,C}$, when the fragment retains all carbon atoms or is limited by the number of fragment ion carbon atoms $\kappa^{F,C}$. This can be described by the set in equation (5.1), in which $\lambda^{P,C}$, $\kappa^{P,C}$, and $\kappa^{F,C}$ act as parameters:

$$TMI_C(\lambda^{P,C}, \kappa^{P,C}, \kappa^{F,C}) = \{\lambda^{F,C} \in \mathbb{N} \mid \lambda^{F,C} \geq \max([\lambda^{P,C} - (\kappa^{P,C} - \kappa^{F,C}), 0]) \wedge \lambda^{F,C} \leq \min([\kappa^{F,C}, \lambda^{P,C}])\} \quad (5.1)$$

The set of fragment labeling states for all MIs of one precursor ion can be combined to yield the total set of mass transitions defining one TMID. Given a pre-specified set of precursor and appendant fragment ions, the total set of mass traces to be extracted can be computed analogously. For fragment ion trace extraction, an extraction width must be specified indicating the symmetric interval centered

on the exact mass of a peak from which all mass peak intensity is retrieved. Smaller values increase selectivity, but reduce sensitivity. This parameter will be 0.001 Da in the following. During the computation of the mass traces, TMIDs are automatically removed from the set of total mass traces when the extraction window around one of its mass traces overlaps with the one of any other included mass trace.

13.2.3 Exemplary Acquisition

This acquisition method was applied to a dilution series of an amino acid standard solution ranging from 50 μM down to 0.1 μM . 50 μl of each dilution was spiked with 50 μl fully ^{13}C -labeled IS, of which 5 μl were injected into the LC-QqTOF system, separated and measured by the acquisition method described above. The Total Ion Count of each measured Q1 window is shown in Figure 3 in the Appendix to this work. Except for L-histidine (HIS) and L-tryptophane (TRP) the chromatography was fairly good, such that at least one fragment ion traces could be integrated per metabolite which yielded coefficients of determination of more than 0.99 along the dilution series. Selected calibration curves are shown in Appendix 10 of this thesis.

13.3 Carbon identity elucidation of CID fragments of primary metabolites

13.3.1 Mixed elemental composition of nominal LC-ESI-MS/MS peaks

Given a working acquisition method, the remaining bottleneck for QTOF MS-based ^{13}C -profiling lies within the structural annotation of fragment ions. Since ^{13}C -isotopomer profiling is ordinarily performed on central carbon metabolism intermediates, the focus of this work lies on collisional fragments of central carbon metabolism intermediates, nucleoside phosphates and amino acids which exceeded a threshold of 3 % intensity relative to the most intense product ion (including the quasi-molecular ion) in a product ion spectrum recorded as described in section 7.8. Each spectrum is provided in Electronic Supplementary Material S7 of Kappelmann et al. [91]. The intensity criterion notwithstanding, all fragments of the study of Rühl et al. [156] were included, for which an accurate mass could be found. Furthermore, fragments were considered that do not satisfy the intensity criterion, but closely neighbor an intense fragment on the m/z axis, what may generate overlapping mass traces in an MRM assay depending on the elemental composition of the neighboring fragments. All considered fragments are listed in Electronic Supplementary Material S1 of Kappelmann et al. [91]. The accurate mass entries from the MassBank server [78] were used, complemented with own experimental accurate mass measurements (see above) to determine the elemental composition of each product ion from all admissible elemental combinations. For details the reader is referred to Electronic Supplementary Material S9 of Kappelmann et al. [91].

Overall, the elemental composition determined by this approach matched the determined elemental composition proposed in the study of Rühl et al. for the vast majority of fragments. A unique elemental composition could be found for all fragments of which the number of carbon atoms could

not be determined in Rühl et al. Noteworthy, discrepancies to the proposed number of carbon atoms in the study of Rühl et al were uncovered for several QqQ mass peaks, which are highlighted in Grey in Electronic Supplementary Material S1 of Kappelmann et al. [91] along with those discrepancies concerning carbon origin (see below). It turned out that four of these QqQ peaks with discordant elemental composition are indeed composed of two distinct accurate mass peaks, which can only be differentiated by accurate mass spectrometry. Therefore, all studied fragments were reviewed for multiple accurate mass peaks and each such instance is highlighted in orange in Electronic Supplementary Material S1 of Kappelmann et al. [91]. Figure 5.2 exemplarily depicts the corresponding peaks from the QqTOF mass spectra of 6PG, GLU and ICIT along with proposed structures for each pair of fragments. It can be clearly seen that no signal is negligible in comparison to the other, such that a QqQ MS-based MRM protocol measuring the mass transitions of one of these fragments would actually measure a superposition of two different carbon backbones. The study of Alves et al. [7] profiled the ^{13}C enrichment of GLU in the m/z 41 fragment ion, shown in Figure 5.2 B. The carbon identity the authors refer to is the one of the $m/z=41.0033$ fragment. The effect on the accuracy of measured fractional abundance cannot be assessed since no direct comparison of NMR multiplet fractions of GLU and measured LC-ESI-MS/MS enrichments in GLU is available.

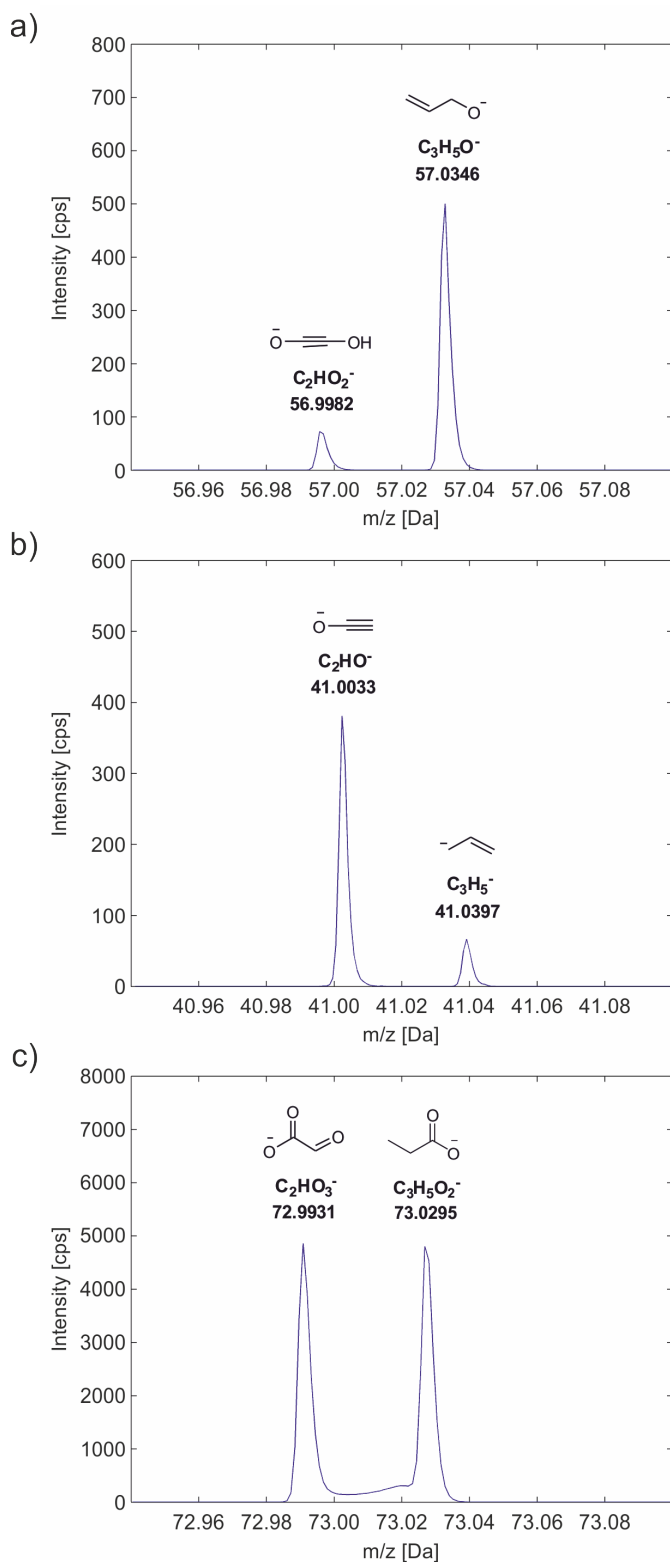


Figure 5.2: Accurate mass peaks of the m/z 57 of 6PG (a), m/z 41 of GLU (b), and m/z 73 of ICIT (c) in negative ionization mode. The product ion spectra were recorded over a CE ramp from -50 eV to -5 eV.

13.3.2 Systematic identification of isotopic interferences in metabolite spectra

The elemental composition of fragments determined in Section 13.3.1 was used to algorithmically compute the mass traces corresponding to carbon labeling states of each fragment with monoisotopic masses of the elements and carbon-13 [186], accounting for the electron mass. With product ion detection at unit mass resolution, the labeling states of neighboring fragments of one metabolite will or will not, depending on the elemental composition, overlap. Such isotopic interferences between labeling states of neighboring fragments can hardly be avoided since the CE parameter can almost never be tuned to solely generate the fragment in question. Electronic Supplementary Material S2 of Kappelmann et al. [91] lists all fragments which suffer from a nominal overlap between tandem mass isotopomers of neighboring fragments and reveals that almost 50 % of all studied fragments (intense and non intense fragments) are affected. Only tandem mass isotopomer detection using high resolution MS instrumentation provides the necessary selectivity to discern tandem mass isotopomers of neighboring fragments of equal nominal mass. The needed mass accuracy to achieve this is determined by the smallest mass difference between any two exact mass traces of equal nominal Q1 and Q3 mass. Therefore, for every such nominal overlap on Q1 and Q3, the exact mass difference between the corresponding mass peaks was computed. The resulting distribution of the mass difference over all such pairs has a median mass difference of 0.0089 Da, whereas its minimum of 0.000264 Da would allow for a maximal mass extraction window width of ± 0.0001 . Since an optimal calibration for a state-of-the-art QqTOF device would result in 0.0001 mass difference between computed and measured masses, this overlap cannot be realistically expected to be resolved. Assuming an error of 0.0004 Da, a total extraction width of 0.001 Da is warranted. Using this best-case mass extraction width, only 2 % of the isotopic interferences would not be resolved. Therefore, high resolution mass spectrometry for product ion detection will prove indispensable to harness the maximal amount of positional ^{13}C -labeling information from CID collisional fragments. Apart from QqTOF MS technology, QOrbitrap MS platforms may serve this purpose. The latter instrument platform was shown to be comparable in terms of quantification accuracy to conventional QqQ MS technology [51], whereas the former one is routinely used in quantitative proteomics [61].

13.3.3 Structural assignment of product ion peaks

In many cases the structure of the product ion alone suffices to deduce its carbon ancestry when orientation-providing heteroatoms like nitrogen and phosphorous or functional groups are present in the fragment ion. Therefore, the structural annotation of product ions is a pivotal step in deducing the positional origin of the product ions carbon atoms. For the structural annotation of product ion peaks the sources of evidence in this study were threefold:

- 1) A structural assignment to mass peaks of known elemental composition was performed by assuming a linear substructure of the precursor ion consistent with the elemental composition (all structures are reported in Electronic Supplementary Material S7 of Kappelmann et al. [91]). Where

possible, already published fragment structures from the studies of Dookeran et al., Rogalewicz et al., and El Aribi et al. were included [43, 155, 49]. For many fragments alternative carbon bond fissures of the precursor ion yield the same fragment structure, resulting in several admissible substructures for one elemental composition. In this case, all alternative structures were exhaustively enumerated, when they resulted in alternative carbon origins (differentiated by the keyword “OR” in Electronic Supplementary Material S7 of Kappelmann et al. [91]). Alternative carbon origins may also arise from intramolecular cyclization reactions since these reactions establish a new bond not present in the precursor ion. When chemically plausible, fragment ion structures were allowed to be derived from intermediates which correspond to intramolecular rearrangement products of the precursor ion. As possible rearrangement reactions the intramolecular formation of (1) anhydrides, (2) esters (lactones), (3) imides and (4) amides were considered. These cyclic intermediates establish a carbon origin not apparent from the precursor ion. In many cases the elemental annotation of product ion spectra yielded direct experimental evidence of the presence of intermediates stemming from such a rearrangement reaction. One striking example of this class represents the m/z 41.9985 (CON^-) ion of ASP in negative ionization mode (see Scheme S-4 in Electronic Supplementary Material S11 of Kappelmann et al. [91]). From the mass spectrum of $[1\text{-}^{13}\text{C}]\text{ASP}$ and $[4\text{-}^{13}\text{C}]\text{ASP}$ it becomes apparent that the carbon atom originates from position 1 and 4 concurrently (see Electronic Supplementary Material S7 of Kappelmann et al. [91]). It appears that this ion is derived from a cyclic intermediate structure, the maleimide anion, apparent at $m/z=96.0091$ ($\text{C}_4\text{H}_2\text{NO}_2^-$) in the mass spectrum of ASP, establishing a link between the α -nitrogen atom and C4 or C1 carbon atoms of ASP. Indeed, the mass spectrum of the $[\text{M}-\text{H}]^-$ ion of maleimide exhibits the m/z 41.9985 as sole intense peak. Analogously, the elemental composition of $\text{C}_2\text{H}_3\text{OS}^+$ for the $m/z=74.9899$ fragment of the $[\text{M}+\text{H}]^+$ ion of methionine clearly exposes this fragment as product of a rearrangement since oxygen and sulfur are connected through two carbon atoms. The mass spectrum of $[5\text{-}^{13}\text{C}]\text{methionine}$ indicates the retention of the *S-methyl* group in this ion, such that it appears likely to be derived of a cyclic intermediate at $m/z=105.0369$ through elimination of ethane (see Electronic Supplementary Material S7 of Kappelmann et al. [91]). This intermediate, in turn, can be derived through carbon monoxide elimination from the cyclic intermediate at $m/z=133.0318$ proposed by Rogalewicz et al. [155]. These clues were used as guidance to the structural annotation of higher mass fragments.

2) In some cases insight into the fragmentation mechanism enabled the elucidation of the carbon origin. For instance, Eichinger et al. [47] reported on the fragmentation mechanism of PYR derivatives involving a CO elimination from the C1 of the PYR backbone. As it turns out, this mechanism allows to assign a structure to all fragment ions derived from α -ketoacids by CO elimination (see fragmentation scheme of αKG in Scheme S-5 in Electronic Supplementary Material S11 of Kappelmann et al. [91]). Furthermore, known fragmentation mechanisms for dicarbonic acids [66] and α -hydroxyl acids [64] were used to aid the structural annotation of OAA, ASP and MAL.

3) The third source of evidence for the structural interpretation of product ions consisted in the use of selectively labeled ^{13}C -standards which helped to delineate isobaric fragment structures for

one elemental composition.

13.3.4 Mixed carbon origin for fragments of sugar phosphates

To elucidate the carbon origin of selected fragment ions whose carbon origin was not apparent from the structural annotation selectively labeled standards were used. From the group of sugar phosphates fragments from F6P, G6P, 6PG and FBP in negative ionization mode were investigated. Such isotopic standards are commercially not available or prohibitively expensive. Therefore, such standards were generated from less expensive ^{13}C -labeled standards by means of *in vitro* enzymatic assays, which were separated using LC. During the elution the mass transitions corresponding to labeling states of certain fragment ions were acquired by means of an MRM assay according to section 7.9. If just one fragmentation reaction were to contribute to the intensity of a given fragment peak, only a single mass trace should bear the total signal intensity (up to isotopic purity) when clearly defined isotopic standards are used. The identity of the mass trace for different isotopic standards on which the total intensity appears allows to deduce the identity of the fragment. When the isotopic purity of such standards is known, the expected intensity distribution across the mass traces can be computed, analogously to the computation of the natural isotopic abundance. This computation assumes knowledge of the neutral loss such that a carbon identity (reflecting itself in the retention or loss of the ^{13}C -label) has to be assumed for each fragment. Comparison of the theoretical distribution to the experimental data allows to assess the correctness of the assumed carbon origin. The expected signal intensity for selectively labeled standards was algorithmically computed using 0.99 as atomic enrichment for the ^{13}C -label and the natural isotopic abundances given in van Winden et al. [181]. The implementation is provided in the form of a zipped MATLAB m-file as Electronic Supplementary Material S8 in Kappelmann et al. [91]. For naturally labeled species this implementation yields the same result as the software from Ramaley et al. [152] for the computation of natural isotopic abundance for MS/MS data.

Table 5.1 lists the result for investigated fragments of sugar phosphates. All signals were acquired with a setting for DP, CE and CXP optimizing the signal intensity for the product ion in question. For each labeled species of the investigated metabolites the experimentally measured fractional FA in the studied fragment is given in boldface type. Each value is accompanied by the theoretical isotopic distribution in italic font. This represents the expected FA which would arise if the assumed carbon origin indicated after the fragment mass was true. The mean standard deviation in FA over replicate injections was 0.01, with a maximum of 0.03. Most importantly, replication is afforded by using different isotopomers of each metabolite. For the m/z 129 ($\text{C}_5\text{H}_5\text{O}_4^-$) fragment of 6PG almost all signal intensity appears on $\text{M1}>\text{m1}$ for $[1\text{-}^{13}\text{C}]\text{6PG}$ as well as $[2\text{-}^{13}\text{C}]\text{6PG}$ and on $\text{M1}>\text{m0}$ for $[6\text{-}^{13}\text{C}]\text{6PG}$. Since the measured FA of each mass trace for all isotopic species agrees well with the expected distribution when this fragment was derived from carbon atoms [1-5], the carbon origin of this fragment can be clearly established. The deduced carbon origin suggests a product ion arising from a single carbon bond break. Any alternative carbon origin would necessitate a product ion structure that re-

Table 5.1: Relative peak areas of individual mass traces for selectively ^{13}C -labeled 6PG, F6P, and FBP. M0, M1, and M2 signify that the quadrupole 1 was set to the mass of the M0, M1, and M2 mass isotopomer of the metabolite in question. m1, m2, and m3 denote that the quadrupole 3 was set to the mass of the m0, m1 and m2 tandem mass isotopomer of the fragment in question. The arising precursor-daughter mass transitions can be found in Electronic Supplementary Material S3 of Kappelmann et al. [91]. Bold face: Measured experimental relative mass trace intensity, Italic face: Theoretical Fractional Abundance arising from the assumed carbon origin of the fragment indicated after the fragment mass. The theoretical abundance was calculated according to the method provided in Electronic Supplementary Material S8 of Kappelmann et al. [91]

	M0		M1		M2	
	m0	m0	m1	m0	m1	m2
Labeled species	Relative signal intensity					
6PG	<i>m/z=129 (Assumed origin: [1-5])</i>					
^{12}C	0.90 (0.91)	0.02 (0.01)	0.05 (0.05)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)
1- ^{13}C	0.00 (0.01)	0.01 (0.00)	0.93 (0.91)	0.00 (0.00)	0.01 (0.01)	0.05 (0.04)
2- ^{13}C	0.01 (0.01)	0.01 (0.00)	0.93 (0.91)	0.00 (0.00)	0.01 (0.01)	0.04 (0.04)
1,2- ^{13}C	0.00 (0.00)	0.00 (0.00)	0.02 (0.02)	0.00 (0.00)	0.01 (0.00)	0.97 (0.92)
6- ^{13}C	0.06 (0.01)	0.88 (0.91)	0.00 (0.00)	0.00 (0.00)	0.05 (0.05)	0.00 (0.00)
F6P	<i>m/z=101 (Assumed origin: [3-6])</i>					
^{12}C	0.93 (0.92)	0.02 (0.02)	0.04 (0.04)	0.01 (0.01)	0.00 (0.00)	0.00 (0.01)
1- ^{13}C	0.04 (0.01)	0.67 (0.92)	0.24 (0.00)	0.01 (0.01)	0.04 (0.04)	0.01 (0.00)
2- ^{13}C	0.05 (0.01)	0.62 (0.92)	0.28 (0.00)	0.01 (0.01)	0.03 (0.04)	0.01 (0.00)
1,2- ^{13}C	0.01 (0.00)	0.04 (0.02)	0.01 (0.00)	0.65 (0.92)	0.05 (0.00)	0.24 (0.00)
6- ^{13}C	0.03 (0.01)	0.29 (0.00)	0.62 (0.92)	0.00 (0.00)	0.03 (0.02)	0.02 (0.03)
F6P	<i>m/z=169 (Assumed origin: [4-6])</i>					
^{12}C	0.92 (0.92)	0.03 (0.03)	0.03 (0.03)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)
1- ^{13}C	0.05 (0.01)	0.89 (0.92)	0.00 (0.00)	0.02 (0.02)	0.03 (0.03)	0.00 (0.00)
2- ^{13}C	0.07 (0.01)	0.87 (0.92)	0.00 (0.00)	0.02 (0.02)	0.03 (0.03)	0.00 (0.00)
1,2- ^{13}C	0.01 (0.00)	0.07 (0.02)	0.00 (0.00)	0.92 (0.92)	0.00 (0.00)	0.00 (0.00)
6- ^{13}C	0.01 (0.01)	0.01 (0.00)	0.92 (0.92)	0.00 (0.00)	0.03 (0.03)	0.02 (0.02)
FBP	<i>m/z=139 (Assumed origin: [5-6])</i>					
^{12}C	0.91 (0.91)	0.05 (0.04)	0.02 (0.02)	0.01 (0.01)	0.00 (0.00)	0.01 (0.01)
1- ^{13}C	0.06 (0.01)	0.87 (0.91)	0.01 (0.00)	0.03 (0.03)	0.02 (0.02)	0.00 (0.00)
2- ^{13}C	0.07 (0.01)	0.86 (0.91)	0.01 (0.00)	0.03 (0.03)	0.02 (0.02)	0.00 (0.00)
6- ^{13}C	0.01 (0.01)	0.03 (0.00)	0.90 (0.91)	0.00 (0.00)	0.04 (0.04)	0.01 (0.01)
FBP	<i>m/z=169 (Assumed origin: [4-6])</i>					
^{12}C	0.91 (0.91)	0.04 (0.03)	0.04 (0.03)	0.01 (0.01)	0.00 (0.00)	0.00 (0.01)
1- ^{13}C	0.06 (0.01)	0.89 (0.91)	0.04 (0.00)	0.00 (0.02)	0.00 (0.03)	0.00 (0.00)
2- ^{13}C	0.07 (0.01)	0.89 (0.91)	0.04 (0.00)	0.00 (0.02)	0.00 (0.03)	0.00 (0.00)
6- ^{13}C	0.02 (0.01)	0.05 (0.00)	0.92 (0.91)	0.02 (0.00)	0.00 (0.03)	0.00 (0.02)

quires two bond breaks, therefore being energetically unfavorable. A quite different picture emerges for the m/z 101 ion of F6P ($\text{C}_4\text{H}_5\text{O}_3^-$): All isotopomers consistently show that the fragment is derived

from positions [1-4] and [3-6] concurrently, although the latter product ion involves two bond breaks assuming a linear precursor ion structure. Heuristic fragmentation rules, favoring parsimonious reactions, and considerations based on standard bond energies would predict the carbon origin [1-4] to be unique. These results show that energetically unfavorable fragmentation routes may still be kinetically competitive. For the m/z 169 ($\text{C}_3\text{H}_6\text{O}_6\text{P}^-$) of F6P the well established and unique carbon origin from carbon atoms [4-6] is confirmed by the labeling pattern of Table 5.1. Since F6P and G6P always fragmented identically on the investigated product ions, only F6P is shown such that these results hold equally true for G6P. A unique origin also holds true for the m/z 139 and m/z 169 ions of FBP, although two potential fragmentation routes lead to these product ions due to two distal phosphate groups. Using the above described isotopic standards, a mixed carbon ancestry for the $m/z=151$ and $m/z=249$ peaks of FBP and the $m/z=85$ fragment of G6P was deduced (data not shown).

13.3.5 Mixed carbon origin for fragments of dicarboxylic acids

Fragments of di- and tricarboxylic acids corresponding to a neutral loss of CO_2 are another representative of the class of fragment structures for which alternative carbon origins exist. Neither fragmentation rules nor combinatorial fragmentation based on standard bond energies allow to infer the carbon ancestry from the mother molecule since for each such a fragment a carbon-carbon bond is broken leading to equal scores for alternative structures [71].

The relative peak areas of different mass traces of the MRM assay for OAA, MAL, αKG , GLU and ASP are given in Table 5.2. Again, all signals were acquired with a setting for DP, CE and CXP optimizing the signal intensity for the product ion in question. The mean standard deviation over replicate injections was 0.01 and is therefore omitted. For the m/z 87 fragment of OAA the total signal intensity of $[1-^{13}\text{C}]\text{OAA}$ appears on $\text{M1}>\text{m1}$ which suggests that C4-decarboxylation is the only CID fragmentation taking place under the parameters tested. This is not a surprising result, given the well-known susceptibility of β -keto acids in liquid solution to decarboxylation which proceeds through a six-membered transition state. Obviously, the same mechanism appears to favor C4-decarboxylation in the gas phase, although the C1-C2 carbon bond is more labile. Thus, the favorability of the transition state makes a rearrangement reaction dominant in comparison to the single cleavage of the more labile bond. The break of the most labile bond is one of many heuristic fragmentation rules commonly employed for structural annotation, which led McCloskey et al. [120] to erroneously assume C1-decarboxylation of OAA, resulting in an inaccurate carbon origin.

In the case of the m/z 71 fragment of MAL the data suggest that two fragmentation reactions take place simultaneously, one generating a fragment stemming from carbon atoms [2-4] and the other one from carbon atoms [1-3]. The m/z 73 fragment of MAL is composed of two carbon atoms, as revealed by the elemental annotation (see Electronic Supplementary Material S1 of Kappelmann et al. [91]), which originate from carbon atom positions [1-2] since $[3-^{13}\text{C}]\text{MAL}$ appears on the $\text{M1}>\text{m0}$ trace. Although chemically similar to OAA, αKG appears to undergo a C1- and C5-decarboxylation simultaneously. The CE energy setting generating this result is the one maximizing the signal intensity

of the m/z 101 fragment. The question arises whether this result holds over the entire domain of CE settings. Figure S-1 in Electronic Supplementary Material S10 of Kappelmann et al. [91] shows that each of the fragmentation reactions corresponding to a C1- or C5-decarboxylation has a different energy threshold and a different CE dependence of its signal intensity, suggesting that both reactions indeed yield chemically different product ions. Over the entire domain of CE a mixture of both species contribute to the m/z 101 peak.

Experimentally, the production of isotopic pure [4- ^{13}C]MAL and [4- ^{13}C]OAA proved difficult, evidenced by the higher relative peak areas on the M1 mass traces corresponding for doubly labeled species and the M0 mass traces for singly labeled species comprising a ^{13}C -label in C4. This effect is most likely due to the isotopic dilution of $\text{H}[^{13}\text{C}]\text{O}_3^-$ in the enzymatic assay resulting from the solution of atmospheric carbon dioxide in the distinct assay components. However, at no point in the running discussion does this effect compromise the conclusions drawn from the data.

For GLU as well as ASP the main signal intensity is recorded on the M1>m1 trace for the $[\text{M-H-44}]^-$ fragments. This observation is in conflict with the heuristically determined carbon origin in the studies of Rühl et al. [156], but agrees with experimental study of Grossert et al. [65], who reported the retention of 1- ^{13}C -label in all product ions of GLU and ASP in negative ionization mode. However, the contribution to the M1>m0 trace is higher than the expected intensity due to natural isotopic abundance and isotopic impurity such that both C1 and C4 or C5 neutral losses appear to occur simultaneously. This result inherently depends on the ionization mode because GLU and ASP in positive ionization mode can be expected to solely undergo a C1-decarboxylation due to the high stability of the immonium ion generated upon CO and H_2O loss [122].

Through the use of the ^{13}C -isotopic standards of this study a mixed carbon ancestry was furthermore deciphered for the $m/z=41$, $m/z=71$, $m/z=86$ peaks of ASP, the $m/z=84$ and $m/z=85$ ions of GLU and for the $m/z=54$, 56, 68, 69, 70 ions of lysine. For CIT it was established that each fragment ion has a symmetrical mixed carbon origin except for fragment ions which are symmetric around the prochiral C3 carbon atom (see below). For ICIT it holds that the vast majority of fragments is of mixed carbon ancestry as are the $m/z=85$ and $m/z=129$ fragments of cis-ACO. Concluding, the competitive fragmentation reactions generating isobaric product ions of alternate carbon origin are a rather common phenomenon in the product ion spectra of central carbon metabolism. Therefore, the use of parsimonious and heuristic rules for carbon identity elucidation is cautioned. Rather, all chemically admissible fragmentation routes should be taken into account to avoid misassigning carbon identity.

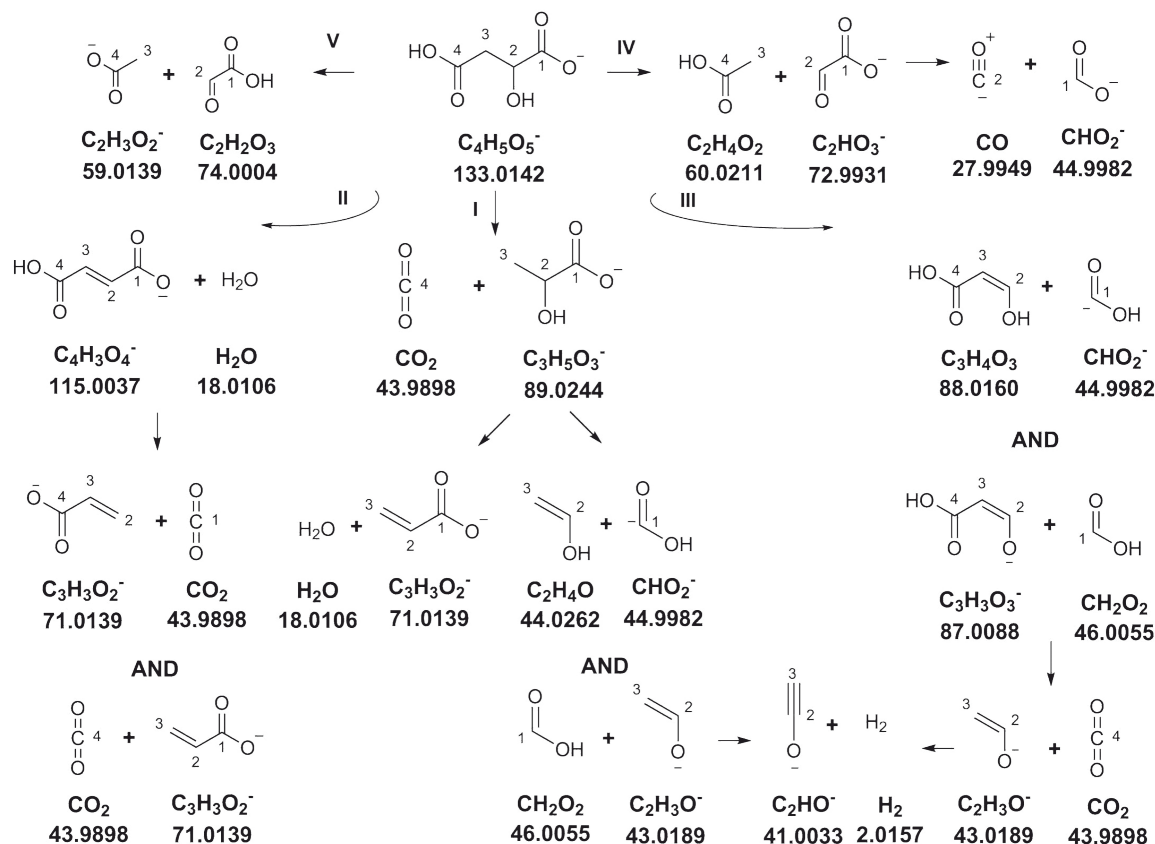
Table 5.2: Relative peak areas of selected mass traces for selectively ¹³C-labeled OAA, MAL, αKG, GLU and ASP. M0, M1, and M2 signify that the quadrupole 1 was set to the mass of the M0, M1, and M2 mass isotopomer of the metabolite in question. m1, m2, and m3 denote that the quadrupole 3 was set to the mass of the m0, m1 and m2 tandem mass isotopomer of the fragment in question. The arising precursor-daughter mass transitions can be found in Electronic Supplementary Material S3 of Kappelmann et al. [91]. Bold face: Measured experimental relative mass trace intensity, Italic face: Theoretical FA arising from the assumed carbon origin of the fragment indicated after the fragment mass. The theoretical abundance was calculated according to the method provided in Electronic Supplementary Material S8 of Kappelmann et al. [91].

	M0	M1		M2	
	m0	m0	m1	m1	m2
Labeled species	Relative signal intensity				
OAA	<i>m/z=87 (Assumed origin: [1-3])</i>				
¹² C	0.94 (0.95)	0.01 (0.01)	0.03 (0.03)	0.00 (0.00)	0.01 (0.01)
1- ¹³ C	0.01 (0.01)	0.01 (0.00)	0.94 (0.95)	0.01 (0.01)	0.03 (0.02)
4- ¹³ C	0.03 (0.01)	0.93 (0.95)	0.00 (0.00)	0.03 (0.03)	0.00 (0.00)
1,4- ¹³ C	0.00 (0.00)	0.01 (0.01)	0.04 (0.01)	0.93 (0.95)	0.00 (0.00)
MAL	<i>m/z=71 (Assumed origin: [1-3])</i>				
¹² C	0.97 (0.95)	0.01 (0.01)	0.02 (0.03)	0.00 (0.00)	0.00 (0.00)
1- ¹³ C	0.01 (0.01)	0.32 (0.00)	0.62 (0.95)	0.02 (0.01)	0.01 (0.02)
4- ¹³ C	0.07 (0.01)	0.57 (0.95)	0.32 (0.00)	0.02 (0.03)	0.01 (0.00)
1,4- ¹³ C	0.00 (0.00)	0.02 (0.01)	0.03 (0.01)	0.92 (0.95)	0.00 (0.00)
3- ¹³ C	0.02 (0.01)	0.01 (0.00)	0.94 (0.95)	0.01 (0.01)	0.02 (0.02)
MAL	<i>m/z=73 (Assumed origin: [1-2])</i>				
¹² C	0.94 (0.95)	0.02 (0.02)	0.02 (0.02)	0.00 (0.00)	0.01 (0.01)
1- ¹³ C	0.01 (0.01)	0.02 (0.00)	0.92 (0.95)	0.02 (0.02)	0.02 (0.01)
4- ¹³ C	0.05 (0.01)	0.92 (0.95)	0.00 (0.00)	0.02 (0.02)	0.00 (0.00)
1,4- ¹³ C	0.00 (0.00)	0.01 (0.01)	0.05 (0.01)	0.92 (0.95)	0.00 (0.00)
3- ¹³ C	0.02 (0.01)	0.95 (0.95)	0.00 (0.00)	0.02 (0.02)	0.00 (0.00)
αKG	<i>m/z=101 (Assumed origin: [1-4])</i>				
¹² C	0.99 (0.94)	0.01 (0.01)	0.00 (0.04)	0.00 (0.00)	0.00 (0.00)
1- ¹³ C	0.00 (0.01)	0.18 (0.00)	0.82 (0.94)	0.00 (0.01)	0.00 (0.03)
GLU	<i>m/z=74 (Assumed origin: [1-2])</i>				
¹² C	0.94 (0.93)	0.03 (0.03)	0.03 (0.02)	0.00 (0.00)	0.00 (0.00)
1- ¹³ C	0.00 (0.01)	0.01 (0.00)	0.94 (0.93)	0.03 (0.03)	0.02 (0.01)
GLU	<i>m/z=102 (Assumed origin: [1-4])</i>				
¹² C	0.92 (0.93)	0.01 (0.01)	0.05 (0.05)	0.00 (0.00)	0.00 (0.00)
1- ¹³ C	0.03 (0.01)	0.09 (0.00)	0.82 (0.94)	0.02 (0.01)	0.04 (0.04)
ASP	<i>m/z=88 (Assumed origin: [1-3])</i>				
¹² C	0.94 (0.94)	0.01 (0.01)	0.04 (0.04)	0.00 (0.00)	0.00 (0.00)
1- ¹³ C	0.01 (0.01)	0.07 (0.00)	0.88 (0.95)	0.01 (0.01)	0.02 (0.03)

13.3.6 Fragmentation pathway of malate

The experimental approaches hitherto applied were complemented by examining the product ion spectra of selected metabolites for generic fragmentation patterns published so far. One such pattern was deduced by Greene et al. [64] for α -hydroxyacids: The negative ions of these metabolites concomitantly yield a m/z 45 ion and a $[M-H-46]^-$ fragment, which the authors showed to correspond to complementary pairs of product ion and neutral loss upon loss of formic acid. The product ion spectrum of MAL exhibits not only those two fragments, i.e. m/z 45 and m/z 87 (see pathway **III** in Scheme 5.1), but also those of lactic acid (m/z 43 and m/z 45). It appears that the lactic acid anion is generated from MAL through C4-decarboxylation, giving rise to fragmentation pathway **I** in Scheme 5.1. The m/z 43 fragment may further eliminate hydrogen at higher collision energies, yielding the m/z 41 ion. Further fragmentation routes comprise a McLafferty-type rearrangement with its archetypal product ion at m/z 59 (pathway **V**) [66] and the elimination of acetic acid (pathway **IV**). The latter one generates the glyoxylic acid anion. The spectrum of this ion, in turn, exhibits just one peak at m/z 45, corresponding to the formic acid anion. Since this anion is again derived from C1, the unique carbon origin of the $m/z=45$ ion is preserved, although various reactions contribute to this peak. Pathway **II** proceeds via the elimination of water generating the symmetric fumaric acid anion. This symmetry explains the mixture of carbon ancestry of the m/z 71 fragment derived from this intermediate by decarboxylation (Table 5.2). However, since the fumarate anion is symmetric, an equal partition of the signal between $M1>m0$ and $M1>m1$ would be expected, which is apparently not the case. Due to the higher signal intensity on the $M1>m1$ mass trace for $[1-^{13}\text{C}]\text{MAL}$, there needs to be an additional labeled contribution to the m/z 71 peak. Elimination of water from the lactic acid anion, containing the $[1-^{13}\text{C}]$ -label, would account for this observation. Indeed, the m/z 71 ion appears in the product ion spectrum of the $[M-H]^-$ ion of lactate, the accurate mass of which matches $\text{C}_3\text{H}_3\text{O}_2^-$. This pathway is in agreement with the elemental composition determined above and the data presented in Table 5.2 and Table 5.3 (see below). It explains the apparent product ion spectrum of MAL, which is given in the Electronic Supplementary Material S7 of Kappelmann et al. [91].

In the case of MAL, it turns out that its product ion spectrum can be interpreted as a superposition of the fragmentation reactions of smaller molecules which, in turn, are derived from the $[M-H]^-$ ion of MAL by simple one-step elimination reactions, i.e. decarboxylation, elimination of water, acetic acid and formic acid. This observation prompted us to ask whether necessarily more complex spectra of larger molecules can also be understood in terms of more basic fragmentation reactions. For this purpose, CIT was chosen as central molecule of the TCA cycle which is amenable to a multitude of fragmentation channels.



Scheme 5.1: Fragmentation pathway of the $[\text{M}-\text{H}]^-$ ion of MAL explaining the apparent product ion spectrum in negative ionization mode. The product ion spectrum can be found in Electronic Supplementary Material S7 of Kappelmann et al. [91]. For each product ion and neutral loss the chemical formula and exact mass are indicated. “AND” signifies that the presence of complementary product ion species of one fragmentation reaction has been confirmed either by them generating a peak in the product ion spectrum or by selectively labeled species.

13.3.7 Fragmentation pathway of citrate

As a first step, the assay of section 7.9 was used to generate $[1-^{13}\text{C}]\text{CIT}$, $[2-^{13}\text{C}]\text{CIT}$, $[3-^{13}\text{C}]\text{CIT}$ and $[6-^{13}\text{C}]\text{CIT}$, corresponding to unlabeled ACCOA reacting with $[4-^{13}\text{C}]\text{OAA}$ and $[3-^{13}\text{C}]\text{OAA}$, $[2-^{13}\text{C}]\text{OAA}$ $[1-^{13}\text{C}]\text{OAA}$, respectively. Furthermore, $[1,6-^{13}\text{C}]\text{CIT}$, $[1,2-^{13}\text{C}]\text{CIT}$ and $[1,3-^{13}\text{C}]\text{CIT}$ were included to afford replication. The acquisition method described in section 7.9 was used to acquire the labeling enrichment in each of the product ions of CIT for the above mentioned selectively labeled species in technical triplicate. To aid the interpretation, an analogous dataset for ICIT as well as cis-ACO was acquired. The ^{13}C -enrichment across all fragments of CIT, ICIT and cis-ACO for the above mentioned isotopomers is shown in Electronic Supplementary Material S6 of Kappelmann et al. [91] together with the standard deviation and expected signal intensity assuming a unique carbon origin. Tracking which carbon atom is retained and which carbon atom is lost, one is able to exclude many structures explaining a given elemental composition. For instance, the $m/z=87.0088$ fragment of CIT ($\text{C}_3\text{H}_3\text{O}_3^-$) admits two chemically different fragments, one corresponding to 2-oxopropanoic acid (i.e. PYR) from

carbon atoms [2,3,6] or [3,4,6], the other being 3-oxopropanoic acid (malonate semialdehyde) from carbon position [1,2,3] or [3,4,5]. Each structure admits two alternative carbon origins, necessitating the differentiation of four different carbon origins. The data in Electronic Supplementary Material S6 of Kappelmann et al. [91]. reveal that this fragment is derived symmetrically from carbon atoms [2,3,6] and [4,3,6]. Thus, its structure can be assumed to be the PYR anion. Analogously, given the labeling data the space of possible structures for the remaining fragment ions can be restricted to the ones given in Figure 5.3. For all compounds for which authentic standards could be obtained down to the mass of 85.0295 a QqTOF product ion spectrum was recorded as described in section 7.8, the mass peaks of which are listed in the column corresponding to the compound in question. Smaller molecules are unlikely to fragment at all or their fragmentation is inconsequential given that the running analysis is restricted to the mass range from 40 to 191. The mass spectrum of CIT exhibits two product ions at $m/z=173.0092$ and $m/z=154.9986$, corresponding to the loss of one and two water molecules, respectively. For aliphatic hydroxylated compounds the loss of water is conventionally assumed to originate from the hydroxyl group [64]. By this token, a racemic mixture of cis- and trans-ACO would contribute to the $m/z=173.0092$ peak. For the second loss of water, only two processes could account: Either it is eliminated from a carboxyl group or an intracellular anhydride is formed. The latter process would generate cis-aconitic anhydride and trans-aconitic anhydride. However, the product ion spectrum of neither was concordant with that of CIT: Both compounds yield one fragment not apparent in the mass spectrum of CIT ($m/z=83.0139$). This observation was taken as evidence against their presence in the fragmentation pathway. Thus, the assumption that the second water molecule is eliminated from a carboxyl group is compelling. Surprisingly, the ^{18}O -enrichment in both fragments for [*hydroxy*- ^{18}O]CIT (shown in Electronic Supplementary Material S6 of Kappelmann et al. [91]) reveals that this reaction sequence accounts for a small share of signal intensity. Instead, the larger part of signal intensity is accounted for by the reverse sequence of initial water elimination from a carboxyl group with subsequent water loss arising from the hydroxyl moiety. The structure contributing to the m/z 173.0092 peak of ICIT even fully retains the hydroxyl group. For the water elimination from carboxyl groups a mechanism was proposed by Grossert et al. [65] in which a ynoate-ion is generated following the elimination of water. However, for these ions no standards are commercially available which is why a fragmentation pattern was assumed for the corresponding ions (see Scheme S-12, S-16 and S-19 in Electronic Supplementary Material S11 of Kappelmann et al. [91]).

citric acid										191.0197
cis-aconitic acid, trans-aconitic acid, 2-hydroxy-2(hydroxyethyl)succinic acid, citric anhydride										173.0092
3-carboxy-5-hydroxypent-2-en-4-ynoic acid, trans-aconitic anhydride , cis-aconitic anhydride										154.9986
citramalic acid										147.0299
oxaloacetic acid or tautomer 2-hydroxyfumaric acid										130.9986
itaconic acid, citraconic acid, mesaconic acid										129.0193
4-hydroxy-2-methylenebut-3-ynoic acid, citraconic anhydride , itaconic anhydride										111.0088
alpha-hydroxyisobutyric acid										103.0401
acetoacetic acid or tautomers 3-hydroxybut-3-enoic acid and 3-hydroxybut-2-enoic acid										101.0244
pyruvic acid or tautomer 2-hydroxyacrylic acid										87.0110
methacrylic acid										85.0295
but-3-en-1-ynolate										67.0189
acetic acid										59.0139
acetone or tautomer prop-1-en-2-ol										57.0346
ethenol										43.0189
prop-1-en										41.0397
ethynol										41.0033

Figure 5.3: Schematic of the fragmentation pathway of CIT using structure names. In each column the fragment ions of the corresponding metabolite appear, ordered by mass. A presumed intermediate structure was assumed to be correct if it only generated product ions also present in the product ion spectrum of CIT itself. The fragments apparent in the product ion spectrum of CIT are listed in the rightmost column. In the leftmost column chemically plausible structures for each product ion CIT are given which are consistent with the labeling data of Electronic Supplementary Material S6 of Kappelmann et al. [91]. The red colored ones are excluded based on other evidence (see main text).

Since no anhydride is formed from cis-ACO and trans-ACO in the fragmentation sequence of CIT, anhydride formation was judged to be unlikely to account for an additional water loss in general. By this token, citric anhydride, citraconic anhydride and itaconic anhydride can be excluded as possible structures for the $m/z=173.0092$ and $m/z=111.0088$ peaks, respectively, although all of their product ions also appear in the mass spectrum of CIT. The $m/z=111.0088$ ion does not appear in the mass spectrum of citraconate nor mesaconate ($m/z=129.0193$), but first appears in the spectrum of cis-ACO, from which the neutral loss is $-H_2O$ and $-CO_2$. Since a sequential loss of CO_2 and H_2O would proceed through citraconate, from which water is obviously not lost, the reverse mechanism of initial water loss with subsequent CO_2 loss must hold true. This mechanism was shown to proceed through a combined loss of H_2O and CO_2 as carbonic acid for glutaric acid [89]. This explains why the $m/z=111.0088$ ion is not generated from citraconate anion ($m/z=129.0193$), but rather from cis-ACO. The mechanism of glutaric acid is directly transferable to cis-ACO since it can be understood as 2-dehydrogenated 3-carboxy-glutaric acid.

Since the m/z 41.0033 peak appears in the mass spectra of many presumed intermediates, it is generated by a multitude of fragmentation channels. One such channel is the elimination of CO_2 and H_2 from PYR ($m/z=87.0088$). PYR for its part is derived from the fragmentation of OAA which arises from CIT through elimination of CH_3COOH . Analogously, beginning with the lowest mass fragments, one is able to recursively construct other fragmentation channels of CIT. The fragments appearing in no mass spectrum other than CIT itself are those corresponding to OAA (m/z 130.9986), citramalate (m/z 147.0299) and ACO and 2-hydroxy-2-(hydroxyethynyl)succinic acid (m/z 173.0092). These compounds arise through the elimination of acetic acid, CO_2 and water. Thus, as in the case of MAL the spectrum of CIT appears as superposition of the fragmentation routes of intermediates arising from these basic eliminations. The fragmentation sequence of the presumed intermediates of CIT is depicted in Electronic Supplementary Material S11 of Kappelmann et al. [91]. These fragmentation sequences also establish the reason as to the symmetrically mixed carbon origin of the majority of CIT fragments: The initial fragmentation reactions yield a symmetric mixture of citramalate (through CO_2 loss from C1 and C5), OAA (through CH_3COOH elimination from the pro-S arm and pro-R arm, respectively) and a racemic mixture of cis/trans-ACO as well as 1 and 5-dehydrated CIT (through water loss) since fragmentation reactions for each pair of fragments are chemically identical. In the following reactions, this symmetry is propagated to smaller fragments. Some fragmentation reactions let each of the two carbon isomers collapse onto one structure with unique carbon origin, centered on the prochiral C3 of CIT, such that the only fragments with unique carbon origin are m/z 103, m/z 85, m/z 57, and m/z 41.0397. However, McCloskey et al. [120] assumed a unique carbon origin for fragments m/z 87, m/z 111 and m/z 147 which is apparently not the case given the data in Electronic Supplementary Material S6 of Kappelmann et al. [91].

It is argued that every mass peak of which the carbon ancestry is established by two fragmentation pathways proceeding through chemically identical intermediates will be of symmetrically mixed carbon ancestry. Reactions proceeding through symmetric fragmentation intermediates are repre-

sentative for this class of reactions. One such fragment is present in the mass spectrum of ICIT: Upon elimination of glyoxylic acid, ICIT generates the SUC anion apparent at $m/z=117.0193$, analogously to the mechanism of glyoxylic acid elimination from MAL. Conspicuously, all ions from the product ion spectrum of SUC also appear in the spectrum of ICIT, distinguishing it from CIT. Since decarboxylation from both terminal carboxyl groups of the SUC ion (originating from carbon atoms [5,6] of ICIT) are chemically identical, the carbon origin of the corresponding peak ($m/z=73.0295$) is indeed symmetrically mixed (see Electronic Supplementary Material S6 of Kappelmann et al. [91]).

In terms of biology, the diastereotopic hydrogen atoms cause the prochiral CIT molecule to react such that a unique carbon atom mapping exists from OAA and ACCOA through CIT to ICIT. For this reason each carbon atom has a unique positional assignment within the CIT molecule. In terms of mass spectrometry, the prochiral side chains of CIT fragment symmetrically, yielding a relative contribution of 0.5 for each of the two alternative carbon origins. Interestingly, as is shown in the Electronic Supplementary Material S12 of Kappelmann et al. [91], this symmetric mixture still contains as much information with regard to positional labeling enrichment as a unique carbon ancestry. This symmetric mixture enabled Alves et al. [7] to derive equations relating the mass fractions from the m/z 67 fragment of CIT to the flux ratio of PCx and PDHC. This ratio could be determined from the CIT mass fractions alone, showing the still informative labeling pattern of symmetric mixed carbon ancestry.

Furthermore, the fragmentation pathways of PYR, MAL, ASP, αKG , OAA, and cis-ACO in negative ionization mode were derived and validated and are provided in Scheme S-2, Scheme S-3, Scheme S-4, Scheme S-5, Scheme S-6, and Scheme S-7, respectively, of the Supplementary Information to Kappelmann et al. [91] with additional discussion. For fragments of PYR and αKG corresponding to CO-elimination the data of this work are fully consistent with the mechanism of CO-elimination proposed by Eichinger et al. [47] for PYR derivatives. The mechanism proposed for ASP is consistent with the MS3 data from the study of Eckersley et al. [45]. The identity of each product ion in the fragmentation pathway was confirmed by recording a product ion spectrum of [1,2,3,4- ^{13}C] αKG , [1- ^{13}C]PYR, [2- ^{13}C]PYR, [3- ^{13}C]PYR, [1- ^{13}C]ASP and [4- ^{13}C]ASP, respectively or by using in-house produced ^{13}C -isotopes. The product ion spectra of αKG , PYR, ASP, MAL, ^{13}C -labeled analogues and pathway intermediates are shown in Electronic Supplementary Material S7 of Kappelmann et al. [91].

13.3.8 Complete isotopomer information for several metabolites

Choi et al. [29] showed that by combining the TMIDs of several fragments the entire isotopomer distribution of the corresponding metabolite may be inferred. For this inference to be feasible, the matrix mapping the isotopomer vector onto the combined tandem mass isotopomer vectors of several fragments has to have full rank. The rank of the mapping matrix for all metabolites was algorithmically computed assuming the TMID of all studied fragments with unique carbon ancestry to be known. For this purpose algorithms inspired by Weitzel et al. [187] were used.

The resulting rank is shown in Electronic Supplementary Material S2 of Kappelmann et al. [91]. Since all studied fragments with unique carbon origin are included, regardless of their intensity, this

rank captures the maximal degree of information with regard to isotopomer resolution afforded by the carbon-backbone fissures generating the studied fragments. The higher the rank, the more informative is the product ion spectrum with respect to full isotopomer resolution; a rank equal to $2^{\kappa^{P,C}}$ where $\kappa^{P,C}$ is the number of carbon atoms in the metabolite signifies that full isotopomer resolution can be achieved.

For the running example of MAL it turns out that the corresponding matrix has full rank. The same holds true for metabolites OAA, ASP and GLX in negative ionization mode and GLY, CYS and ASP in positive ionization mode. In case of MAL, however, Electronic Supplementary Material S2 of Kappelmann et al. [91] also reveals that many mass transitions overlap. Therefore, only accurate mass spectrometry will enable to harvest the full information content contained in the CID product ion spectrum. To demonstrate this contention, an LC-ESI-QqTOF-based approach was followed to acquire the entire product ion spectrum of mass isotopomers of MAL. For this purpose, an equimolar mixture of isotopomers MAL₁₀₀₀, MAL₀₁₀₀, and MAL₀₀₁₀ was produced using the Pyruvate carboxylase assay described in the Experimental section. The resulting relative signal intensities are given in Table 5.3, whereas the M1 product ion spectrum of MAL and the ion chromatogram are given in Electronic Supplementary Material S13 of Kappelmann et al. [91]. Obviously, the fractions represent the expected signal partition according to how many of the labeled atoms are present in the corresponding fragment, yielding either 0.33, 0.66 or 1 (up to isotopic purity). The found accurate mass allows an unambiguous identification of different fragment mass isotopomers in all cases, demonstrating that both good quantification and mass accuracy can be achieved with QqTOF MS technology.

Table 5.3: Relative signal intensities of CID product ions of the M0, M1 and M2 mass isotopomers of MAL. A horizontal line indicates that the corresponding labeling state is not admissible, whereas '<LOD' (limit of detection) means that no mass peak could be integrated.

M0 (Q1=133) ^a		M1 (Q1=134)				M2 (Q1=135)					
m0 ^b		m0		m1		m0		m1		m2	
<i>m/z</i> found ^c	relative Int. ^d	<i>m/z</i> found	relative Int.	<i>m/z</i> found	relative Int.	<i>m/z</i> found	relative Int.	<i>m/z</i> found	relative Int.	<i>m/z</i> found	relative Int.
41.0035	0.01	41.0034	0.36	42.0067	0.62	<LOD	<LOD	42.0072	0.02	<LOD	<LOD
43.0193	0.01	43.0192	0.33	44.0227	0.64	43.0192	0	44.0226	0.02	45.0262	0
44.9988	0.01	44.9984	0.63	46.0019	0.36	44.9987	0	46.0019	0	—	—
59.0146	0.03	59.0143	0.63	60.0178	0.30	59.0146	0.02	60.0180	0.02	61.0200	0
72.9941	0.01	72.9939	0.31	73.9973	0.65	72.9939	0	73.9972	0.02	75.0008	0.01
87.0099	0.01	87.0097	0.28	88.0129	0.68	—	—	88.0131	0.01	89.0167	0.01
89.0254	0.01	89.0254	0.01	90.0289	0.96	—	—	90.0288	0.01	91.0322	0.02
115.0049	0.01	—	—	116.0091	0.96	—	—	—	—	117.0119	0.03

^a M0, M1 and M2 denote the mass isotopomers of MAL. The corresponding quadrupole mass filter is indicated. ^b m0, m1 and m2 denote unlabeled, singly labeled and doubly labeled product ions, respectively. ^c The exact masses are the measured exact masses in this specific experiment. ^d The relative intensities were not corrected for natural isotope abundance.

13.3.9 Statistical information gain of positionally resolved labeling information

Clearly, in the absence of systematic errors a greater number of measured quantities directly translates into improved statistical quality of derived quantities. The number of measured mass traces for an arbitrary TMID is given as $(\kappa^{\text{P,C}} - \kappa^{\text{F,C}} + 1)(\kappa^{\text{F,C}} + 1)$, where $\kappa^{\text{P,C}}$ is the number of carbon atoms in the precursor and $\kappa^{\text{F,C}}$ is the number of carbon atoms in the fragment [28]. Thus, for $0 < \kappa^{\text{F,C}} < \kappa^{\text{P,C}}$ this quantity is always greater than $\kappa^{\text{P,C}} + 1$, the number of measurements contained in a MID. On the other hand, it may be noted that the additional measurements in a TMID necessarily arise from a different mass spectrometric acquisition mode, i.e. MS/MS acquisition as opposed to MS acquisition sufficient for MID acquisition. Due to the loss of signal intensity inherent in MS/MS acquisition, the TMID may be less precise than the corresponding MID of the precursor ion in question. Therefore, a gain in statistical quality from positionally resolved labeling information is not guaranteed.

To analyze the extent to which flux confidence may increase, the central metabolism network of *C. glutamicum* ΔME was used in conjunction with the measurement scenario described in the Supplementary Information to Kappelmann et al. [90]. This measurement configuration is taken as base scenario and encompasses mass isotopomer information for sugar phosphates and organic acids and the positional resolution afforded by the immonium ion of most amino acids in positive ionization mode and is given in Electronic Supplementary Material S4 of Kappelmann et al. [91]. For comparison, this measurement model was extended by positional labeling information from fragment ions of which the carbon identity was identified or validated in this study. The complete list of added labeling information is listed in Electronic Supplementary Material S5 of Kappelmann et al. [91]. For all simulations 13CFLUX2 software suite [188] was used. Essentially, 1000 random label distributions were generated from the network, for each of which the associated flux standard deviations were evaluated for the two measurement scenarios. Figure 5.4 shows the mean log2 of the relative flux standard deviations computed from the extended measurement model in comparison to the one computed from the base scenario. This computation was once performed using an error of 0.01 for the MID and 0.05 for the TMID (black bars) and once assuming an equal measurement error of 0.01 for the MID and TMID (grey bars). For the latter scenario, it can be seen that an up to 20-fold increase in flux precision can be achieved for selected fluxes and that on average the standard deviation is half as high as in the base measurement scenario. Even for the worst case scenario of 0.05 as TMID precision, a significant information gain can be accomplished.

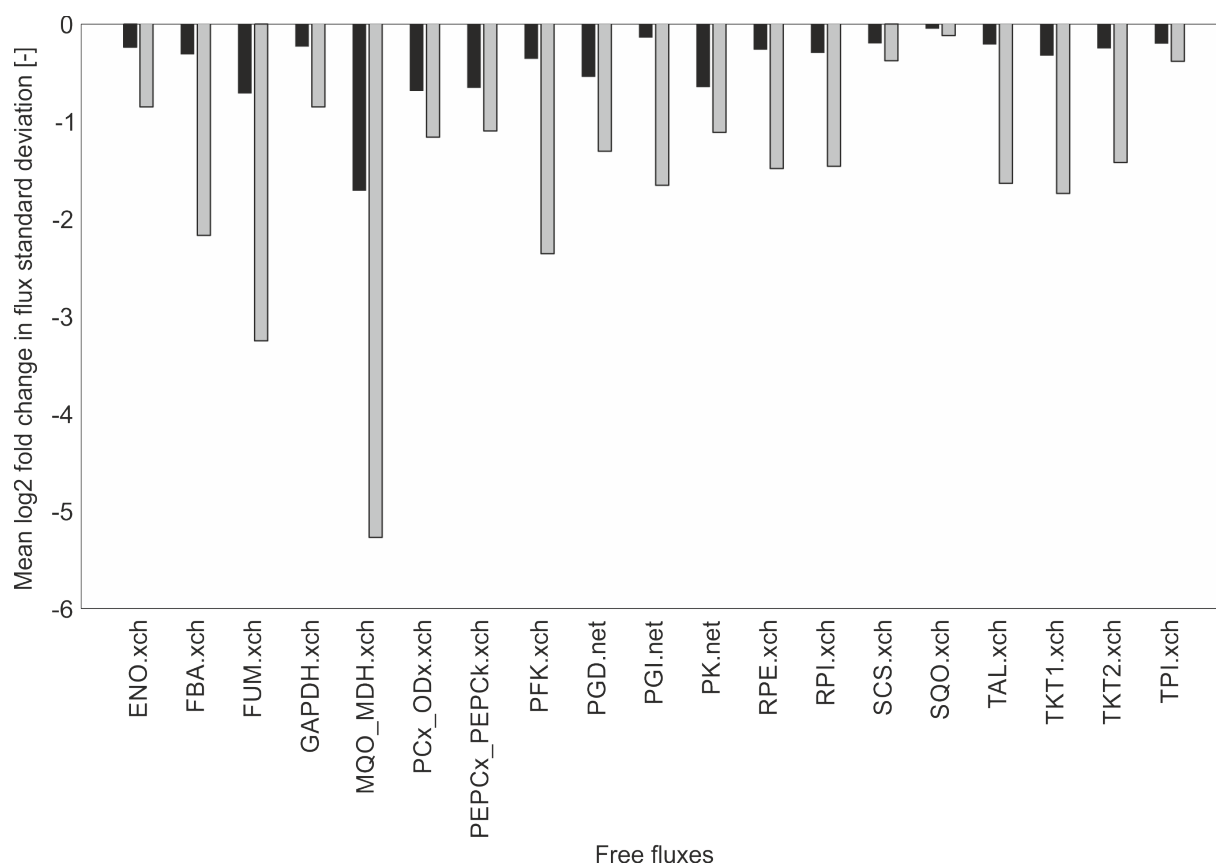


Figure 5.4: Relative flux standard deviations for all free fluxes in comparison to the measurement scenario described in Kappelmann et al. [90]. Black bars represent the measurement scenario of higher acquisition precision for MID than for TMID, whereas grey bars represent the scenario of equal measurement precision for MID and TMID. The flux nomenclature can be found in Kappelmann et al. [90].

14 QqTOF MS technology for simultaneous ^{13}C - and ^{15}N -isotope enrichment profiling

14.1 Motivation

Up to this point the capabilities of QqTOF MS technology for increased positional ^{13}C -isotopomer resolution have been investigated. Tracer studies have also been performed using other stable isotopes; the ^{15}N -isotope of nitrogen featuring most prominently amongst them. Although the isotopically stationary ^{15}N -distribution is less informative than the isotopically stationary ^{13}C -distribution, ^{15}N can be effectively used for kinetic flux profiling [204]. Kinetic flux profiling consists in tracking the time-dependent incorporation of a tracer into central metabolism intermediates upon its introduction into the growth medium [132]. If both isotopes cannot be analytically separated, two separate experiments are needed to independently generate the ^{15}N -label distribution and the stationary ^{13}C -incorporation. If ^{13}C - and ^{15}N -isotopes could be distinguished by mass spectrometry, on the other hand, one experiment could generate both the ^{15}N - and ^{13}C -label distribution and be analyzed in one single acquisition.

Encouragingly, ^{15}N - and ^{13}C -isotopes have different *relative isotopic mass defects*. This term denotes the difference of the monoisotopic mass of an element and the exact mass of its A+1 isotope minus one [175]. Whereas ^{13}C has a positive relative mass defect of 0.00335 u, ^{15}N has a negative relative mass defect of -0.0029 u¹. Such an isotopic mass defect cannot be directly resolved with either QqQ MS or QqTOF MS devices, such that ^{13}C - and ^{15}N -MI species would be aliased. In fact, only Fourier transform ion cyclotron resonance or Orbitrap mass analyzers with mass resolving power of up to 100,000 (FWHM) are able to discriminate ^{13}C and ^{15}N -isotopes. However, these devices are not only expensive, but also entail significantly longer acquisition times for product ion spectrum acquisition [62]. Furthermore, the linearity of their signal response has been put in doubt [139].

Bar a direct resolution of ^{15}N and ^{13}C -MI species in the m/z domain on QqTOF MS devices, the following section explores two indirect approaches to deduce relative abundances of both MI species. The first approach relies on CID fragmentation to separate co-isolated ^{15}N - and ^{13}C -MI species. For reasons explained below, this approach is not applicable to all amino acids. This fact motivated the second approach which relies on a deconvolution of the TOF product ion spectrum.

14.2 Problem statement

In samples from ordinary ^{13}C -labeling experiments, a direct relationship between ^{13}C -MI abundance and signal on the corresponding Q1 mass trace exists since just one MI species is isolated in each Q1 window. When ^{15}N -tracers are introduced alongside ^{13}C -tracers into the metabolism, various MI species collapse onto one Q1 ion stream during measurement due to the nominal mass resolution

¹For the example of carbon-13 the relative isotopic mass defect, denoted by $R\Delta m_{(A+1)}$, is defined as $R\Delta m_{(A+1)13\text{C}} = m_a(^{13}\text{C}) - m_a(^{12}\text{C}) - 1$, where m_a represents the exact mass of an isotope [175].

of quadrupoles. Each MI species differs in the number of labeled nitrogen and carbon atoms, $\lambda^{P,C}$ and $\lambda^{P,N}$, respectively. Hence, one pair $(\lambda^{P,C}, \lambda^{P,N})$ of labeled carbon and nitrogen atoms identifies a precursor ion MI. All pairs of which the sum of labeled nitrogen and carbon atoms is equal to the mass shift k of the Q1 isolation window in question are co-isolated. This set of distinct co-isolated MI species for any given mass shift k is therefore described by equation (5.2):

$$MI(k, \kappa^{P,C}, \kappa^{P,N}) = \{(\lambda^{P,C}, \lambda^{P,N}) \in \mathbb{N} \times \mathbb{N} \mid \lambda^{P,C} + \lambda^{P,N} = k \wedge \lambda^{P,C} \leq \kappa^{P,C} \wedge \lambda^{P,N} \leq \kappa^{P,N}\} \quad (5.2)$$

Thus, the set $MI(k, \kappa^{P,C}, \kappa^{P,N})$ contains all pairs of labeled carbon and nitrogen atoms that add up to the considered total mass shift k and are admissible given the elemental composition of the precursor ion. For the example of ASP, two MI species are co-isolated in the Q1 isolation window corresponding to $k=1$, i.e. singly-labeled $[^{13}\text{C}]\text{ASP}$ and $[^{15}\text{N}]\text{ASP}$.

14.3 Analytical separation of ^{13}C - and ^{15}N -MI by means of CID

14.3.1 Fragment ion and precursor ion selection

The section 5 illustrated how the product ions of CID reactions which eliminate part of the carbon backbone can be used to positionally resolve the ^{13}C -label of primary metabolites. When the first carbon atom, for instance, is eliminated during CID, co-isolated ^{13}C -isotopomers bearing a label in the first position will collapse onto another mass peak as isotopomers that are not labeled in the first position. Hence, the Q1 ion stream is partitioned into distinct fragment ion traces depending on the label position. When the pool of ionized analyte molecules is comprised of ^{13}C - as well as ^{15}N -isotopomers, additional isotopomers bearing a ^{15}N label are co-isolated in each Q1 window. By analogy to the carbon backbone fragmentation, the latter will not contribute to a labeled mass trace of a fragment ion that does not bear any nitrogen atoms. When such a fragment ion additionally retains a carbon atoms of the precursor ion, all ^{13}C -isotopomers are mapped to a labeled mass trace. Thus, the Q1 ion stream is partitioned into ^{13}C -isotopomers and ^{15}N -isotopomers. Such a fragment for the $[\text{M}-\text{H}]^-$ ion of ASP is shown in Figure 5.5. While this fragment does not contain any nitrogen atoms, it retains all carbon atoms of the precursor ion. It can be seen that precursor ion species of total mass shift k bearing $\lambda^{P,N}$ labeled nitrogen atoms are mapped on the ion trace of this fragment corresponding to a $k - \lambda^{P,N} = \lambda^{P,C}$ mass shift. On the other hand, all co-isolated precursor ion species in the same Q1-isolation window with only ^{13}C -label will necessarily contribute to the fragment ion trace corresponding to the mass shift k since all carbon atoms are retained in the fragment.

Because each co-isolated MI species in every Q1 window has one specific mass trace to which no other precursor ion species contributes, the relative intensity of the k times ^{13}C -labeled fragment ion trace and the $(k - \lambda^{P,N})$ times ^{13}C -labeled mass trace directly reflects the relative abundance of the co-isolated ^{13}C - and ^{15}N -MI species in the current Q1 window. In every Q1 isolation window an unique precursor-product ion transition exists such that the entire ^{15}N -MID and ^{13}C -MID can be

reconstructed. Clearly, if the fragment ion does not retain all carbon atoms, some ^{13}C -MI species will also contribute to the unlabeled fragment ion trace of the corresponding product ion.

Figure 5.5 also shows the extension of the canonical designation of MI species (see Figure 5.1) to the case, where two elements in one metabolite are labeled. The designator of each MI is shown to the right it and is composed of a capital M followed by the number of labeled carbon and nitrogen atoms in brackets. Analogously, a TMI will generally be denoted as $m[\lambda^{F,C}, \lambda^{F,N}]$, where $\lambda^{F,C}$ and $\lambda^{F,N}$ denote the number of labeled carbon and nitrogen atoms in the considered fragment, respectively. The notation for mass traces, however, will not distinguish MI species since the first stage of quadrupole-based mass analysis aliases all precursor MI species. Consequently, a mass trace designator has the following general format: $M[k] > m[\lambda^{F,C}, \lambda^{F,N}]$.

With this notation at hand, one can derive the equations which reconstruct the ^{13}C -MID and ^{15}N -MID from the measured ^{13}C -enrichment. Expressed in the notation from above, the ^{13}C -MID corresponds to a vector of relative intensities of the MI species $M[\lambda^{P,C}, 0]$, where $\lambda^{P,C}$ goes from 0 to $\kappa^{P,C}$. Equation (5.3) computes the relative abundance of each ^{13}C -mass isotopomer $M[\lambda^{P,C}, 0]$ from the FA of selected mass traces. The FA is the intensity of a mass trace relative to the summed intensity of all mass traces corresponding to the labeling states of one metabolite.

$$M[\lambda^{P,C}] = \sum_{k=\lambda^{P,C}}^{\kappa^{P,C} + \kappa^{P,N}} \text{FA}(M[k] > m[\lambda^{P,C}, 0]) \quad (5.3)$$

where $\kappa^{P,C} + \kappa^{P,N}$ is the total number of mass shifts for the metabolite in question. The relative abundance of a given ^{15}N -MI can be calculated according to equation (5.4):

$$M[\lambda^{P,N}] = \sum_{k=\lambda^{P,N}}^{\kappa^{P,C} + \kappa^{P,N}} \text{FA}(M[k] > m[k - \lambda^{P,N}, 0]) \quad (5.4)$$

Despite the simplicity of this concept, there are no reports which harvest this property of a CID fragmentation for analytical discrimination of ^{13}C -MI and ^{15}N -isotopes. This is thus the first instance, where ^{13}C - and ^{15}N -isotopes of a metabolite are separated without resolving them in the m/z domain.

The applicability of this concept crucially depends on the presence of intense ions in the CID product ion spectra of amino acids which retain all carbon but no nitrogen atoms. Assuming that mostly even-electron neutrals are lost during CID, the loss of NH_3 alone or combined with H_2O would generate such ions. For amino acids with thiol and hydroxyl groups an additional loss of hydrogen sulfide and water, respectively, is possible. This set of admissible fragment ions for amino acids is listed in Table 5.4, followed by the exact mass of the resulting mass peak. The QqTOF MS product ion spectra of the $[\text{M}+\text{H}]^+$ ions of amino acids recorded over a CE ramp of +5 to +50 eV were searched for these mass peaks. Quantification of their relative abundance with respect to the most intense ion in the spectrum yields the relative intensities shown on the right to each mass peak. It is apparent that the $[\text{M}+\text{H}]^+$ ions of amino acids very rarely eliminate ammonia alone or in combination with other neutral species. Only the CID fragmentation of MET and TYR generate sufficiently intense product

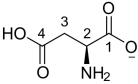
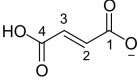
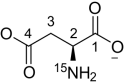
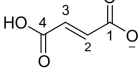
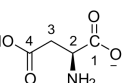
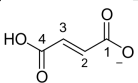
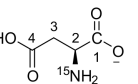
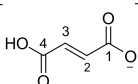
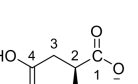
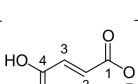
	Co-isolated precursor mass isotopomers	Fragment labeling states for the m/z 115.0037 ion	Corresponding neutral loss	Mass trace designator
Q1=132	 $\text{C}_4\text{H}_6\text{NO}_4^-$ 132.0302	 $\text{C}_4\text{H}_3\text{O}_4^-$ m[0,0]	NH_3 H_3N 17.0265	M[0]>m[0,0]
Q1=133	 $\text{C}_4\text{H}_6^{15}\text{NO}_4^-$ 133.0273	 $\text{C}_4\text{H}_3\text{O}_4^-$ m[0,0]	$^{15}\text{NH}_3$ H_3^{15}N 18.0236	M[1]>m[0,0]
	 $\text{C}_3^{13}\text{CH}_6\text{NO}_4^-$ 133.0336	 $\text{C}_3^{13}\text{CH}_3\text{O}_4^-$ m[1,0]	NH_3 H_3N 17.0265	M[1]>m[1,0]
Q1=134	 $\text{C}_3^{13}\text{CH}_6^{15}\text{NO}_4^-$ 134.0306	 $\text{C}_3^{13}\text{CH}_3\text{O}_4^-$ m[1,0]	$^{15}\text{NH}_3$ H_3^{15}N 18.0236	M[2]>m[1,0]
	 $\text{C}_2^{13}\text{C}_2\text{H}_6\text{NO}_4^-$ 134.0369	 $\text{C}_2^{13}\text{C}_2\text{H}_3\text{O}_4^-$ m[2,0]	NH_3 H_3N 17.0265	M[2]>m[2,0]

Figure 5.5: All possible carbon and nitrogen labeling states for the $[\text{M}-\text{H}]^-$ ion of ASP up to a mass shift of $k = 2$. For each precursor MI the arising labeling state in the $[\text{M}-\text{NH}_3-\text{H}]^-$ fragment ion is shown. A bracket indicates the labeling state of a structure without specifying its position.

ions conforming to the above criteria. Table 5.4 does not contain the amino acids TRP, citrulline, ARG or HIS. From the structure of these amino acids it becomes clear that no chemically plausible fragmentation exists which would eliminate all nitrogen atoms, while leaving the carbon backbone intact.

Although the positive ionization is the most sensitive for amino acids [148], this metabolite class will also easily ionize in negative ionization mode. The fragmentation of the $[\text{M}+\text{H}]^+$ and $[\text{M}-\text{H}]^-$ ions of one metabolite can be dramatically different as the elucidation of product ion structures for the $[\text{M}+\text{H}]^+$ and $[\text{M}-\text{H}]^-$ ions of GLU and ASP in the previous chapter illustrated. Thus, in negative ionization mode some amino acids may eliminate ammonia, the loss of which is unfavorable under positive ionization mode. For this reason the product ion spectra of the $[\text{M}-\text{H}]^-$ ions of amino acids were acquired over an CE-ramp from -5 to -50 eV and screened for the mass peaks corresponding to the neutral losses from above. The relative intensities of the admissible fragment ions is shown to the right of the corresponding exact peak mass. It appears that for some amino acids the product ion corresponding to a loss of ammonia is more favored when it is negatively charged. This is the case for ASP, for which the $[\text{M}-\text{NH}_3-\text{H}]^-$ ion accounts for 7 % relative intensity, whereas the corresponding $[\text{M}-\text{NH}_3+\text{H}]^+$ product ion is completely absent from the respective spectrum. However, Table 5.4 also reveals that positive and negative ionization mode combined still do not yield an intense ion for all amino acids.

It is well known that the ESI process can generate stable adduct ion species with alkali metals, when these are present in the solution introduced into the ESI source. Since the charge carrier in these ions is no longer the metabolite, but the alkali metal cation, only charge-remote fragmentation reactions can lead to a fissure of the covalent bonds of the precursor ion. The resulting spectra are quite distinct from the ones of the corresponding $[\text{M}+\text{H}]^+$ ions and may contain the product ions that are targeted in this work.

Therefore, the $[\text{M}+\text{Na}]^+$, $[\text{M}+\text{K}]^+$ and $[\text{M}+\text{Li}]^+$ adducts of selected amino acid were screened for favorable fragmentation reactions. However, the $[\text{M}+\text{Na}]^+$ and $[\text{M}+\text{K}]^+$ adducts showed little to no fragmentation apart from the release of the alkali metal cation. Only the $[\text{M}+\text{Li}]^+$ adducts fragmented favorably. The relative intensities of the product ion peaks corresponding to the allowed neutral loss combinations of the $[\text{M}+\text{Li}]^+$ ions are shown in the rightmost column of Table 5.4. Apparently, for some amino acids the searched fragments are an order of magnitude more abundant in comparison to either positive and negative ionization mode. However, the relative intensity with respect to the most intense ion is still not sufficient for most amino acids. Therefore, special tuning of the ion optics and collision cell parameters need to be performed to increase the yield of the fragment ions targeted in this study. Furthermore, derivatization strategies of the amino groups to render them better leaving groups may be explored as a way to tailor the CID reaction of each amino acid such that the desired neutral losses occur in high yield.

Table 5.4: Relative intensities of product ions of amino acids corresponding to neutral loss combinations that leave the carbon backbone intact, but eliminate all nitrogen atoms. The intensity was quantified relative to the most intense product ion over a CE ramp of ± 5 -50 eV.

Amino acid	Neutral loss	Precursor ion					
		$[\text{M}+\text{H}]^+$		$[\text{M}-\text{H}]^-$		$[\text{M}+\text{Li}]^{+a}$	
		peak <i>m/z</i>	rel. Intensity	peak <i>m/z</i>	rel. Intensity	peak <i>m/z</i>	rel. Intensity
ALA	$-\text{NH}_3$	73.0284	0	71.0139	1.71e-03	79.0336	3.76e-03
	$-\text{NH}_3\text{-H}_2\text{O}$	55.0178	6.17e-04	53.0033	0	61.0260	0
ASN	-2NH_3	99.0077	3.82e-03	96.9931	0	105.0158	5.06e-03
	$-2\text{NH}_3\text{-H}_2\text{O}$	80.9971	0	— ^b	—	87.0053	0
ASP	$-\text{NH}_3$	117.0182	0	115.0037	7.02e-02	123.0264	3.84e-03
	$-\text{NH}_3\text{-H}_2\text{O}$	99.0077	4.23e-04	96.9931	3.76e-05	105.0158	7.36e-04
CYS	$-\text{NH}_3$	105.0005	3.19e-02	102.9859	0	111.0087	3.66e-01
	$-\text{NH}_3\text{-H}_2\text{S}$	71.0128	0	68.9982	0	77.0209	0
	$-\text{NH}_3\text{-H}_2\text{O}$	86.9899	3.08e-02	84.9754	0	92.9981	0

Table 5.4 Continued: Relative intensities of product ions of amino acids corresponding to neutral loss combinations that leave the carbon backbone intact but eliminate all nitrogen atoms.

GLU	-NH ₃	131.0339	4.91e-03	129.0193	1.12e-03	137.0421	4.64e-02
	-NH ₃ -H ₂ O	113.0233	4.11e-04	111.0088	0	119.0315	0
	-NH ₃ -2H ₂ O	95.0128	1.23e-04	92.9982	0	101.0209	0
GLN	-2NH ₃	113.0233	4.63e-04	111.0088	0	119.0315	7.29e-04
	-2NH ₃ -H ₂ O	95.0128	1.29e-03	92.9982	0	101.0209	0
GLY	-NH ₃	59.0128	6.36e-04	56.9982 ^c	0	65.0209 ^c	3.18e-04
	-NH ₃ -H ₂ O	41.0022	0	— ^b	—	— ^b	—
HOM	-NH ₃	103.0390	0	101.0244	4.70e-04	109.0471	3.51e-03
	-NH ₃ -H ₂ O	85.0284	0	83.0139	0	91.0366	1.05e-04
	-NH ₃ -2H ₂ O	67.0178	0	65.0033	0	73.0260	0
ISOLEU	-NH ₃	115.0754	1.27e-04	113.0608	0	121.0835	1.23e-02
	-NH ₃ -H ₂ O	97.0648	2.65e-04	95.0502	0	103.0730	0
LEU	-NH ₃	115.0754	1.52e-05	113.0608	5.95e-04	121.0835	2.55e-02
	-NH ₃ -H ₂ O	97.0648	1.00e-04	95.0502	0	103.0730	2.86e-05
LYS	-2NH ₃	113.0597	4.17e-04	111.0452	0	119.0679	2.96e-03
	-2NH ₃ -H ₂ O	95.0491	1.76e-03	93.0346	0	101.0573	5.55e-05
MET	-NH ₃	133.0318	5.44e-01	131.0172	0	139.0400	4.69e-02
	-NH ₃ -H ₂ O	115.0212	8-79e-04	113.0067	0	121.03	0
PHE	-NH ₃	149.0597	1.58e-02	147.0452	1.95e-01	155.0679	1.13e-02
	-NH ₃ -H ₂ O	131.0491	1.65e-02	129.0346	0	137.0573	0
PRO	-NH ₃	99.0441	0	97.0295	0	105.0522	5.58e-05
	-NH ₃ -H ₂ O	81.0335	0	79.0189	0	87.0417	0
SER	-NH ₃	89.0233	0	87.0088	0	95.0315	4.12e-04
	-NH ₃ -H ₂ O	71.0128	0	68.9982	0	77.0209	0
	-NH ₃ -2H ₂ O	53.0022	0	— ^b	—	59.0104 ^c	0
THR	-NH ₃	103.0390	0	101.0244	0	109.0471	0
	-NH ₃ -H ₂ O	85.0284	7.01e-04	83.0139	0	91.0366	0
	-NH ₃ -2H ₂ O	67.0178	0	65.0033	0	73.0260	0
TYR	-NH ₃	165.0546	4.71e-01	163.0401	1.46e-01	171.0628	1.21e-02
	-NH ₃ -H ₂ O	147.0441	9.37e-02	145.0295	0	153.05	0
VAL	-NH ₃	101.0597	3.84e-05	99.0452	0	107.0679	7.62e-03

Table 5.4 Continued: Relative intensities of product ions of amino acids corresponding to neutral loss combinations that leave the carbon backbone intact but eliminate all nitrogen atoms.

-NH ₃ -H ₂ O	83.0491	1.68e-04	81.0346	0	89.0573	0
------------------------------------	---------	----------	---------	---	---------	---

^a The product ion spectrum of the ^7Li - ion adduct was considered.

^b The product ion is not admissible given the elemental composition of the precursor ion.

^c The corresponding product ion is a double radical ion species.

It is clear that the above presented approach is accurate because it analytically fully separates all ^{13}C - and ^{15}N -MI species co-isolated in one Q1 isolation window. The experimental validation of this property will be provided using the running example of ASP in section 14.5. As a downside, this approach is not applicable to amino acids with two nitrogen atoms occurring in the form of secondary amines, like in HIS. For these structures no plausible fragmentation can be derived that eliminates all nitrogen atoms while leaving the carbon backbone intact. In an effort to make these amino acids also amenable to analytical separation of their ^{13}C - and ^{15}N -MI species, a different method was devised.

14.4 Spectral deconvolution of reflectron TOF mass peaks for ^{13}C - and ^{15}N -MI separation

14.4.1 General approach

As outlined in the introduction of this chapter, the resolution of a TOF detector will not suffice to discern ^{13}C - and ^{15}N -isotopes. The QqTOF MS device employed in this study is specified to meet a $\text{RP}_{\text{FWHM}} \geq 25,000$ for the scan of m/z 186 [8]. Empirically, this value declines towards lower m/z values. Notwithstanding, a RP_{FWHM} of 20,000 can be assumed for the mass peak of m/z 135, corresponding to the singly labeled $[\text{M}+\text{H}]^+$ ion of ASP. In a ^{13}C - and ^{15}N -colabeling experiment two isotopic species would be isolated in the corresponding Q1-window, i.e. singly-labeled $[^{13}\text{C}]\text{ASP}$ and $[^{15}\text{N}]\text{ASP}$. Based on the RP_{FWHM} of 20,000, each isotopic peak can be expected to be 0.0067 Da wide at half maximal height. Since this width is in the range of the mass difference of both species (0.0063 Da), their mass peaks cannot be realistically expected to be resolved. Indeed, an injection of an equimolar mixture of $[4\text{-}^{13}\text{C}]\text{ASP}$ and $[^{15}\text{N}]\text{ASP}$ does not yield two distinct peaks, but rather a single peak at an accurate mass of m/z 135.0456. However, an injection of only $[4\text{-}^{13}\text{C}]\text{ASP}$ yields an accurate mass of m/z 135.0485, whereas a solution of 100 % $[^{15}\text{N}]\text{ASP}$ yields a peak at m/z 135.0420². Therefore, the mass peaks in a TOF spectrum still bear information with respect to the relative contribution of each species.

The above described shift in found accurate mass for a mass peak composed of two isobaric species has been described earlier [123]. The authors proposed to understand the observed peak

²The indicated masses are the accurate masses from this specific experiment. The exact masses are m/z 135.0481 and m/z 135.0418 for $[4\text{-}^{13}\text{C}]\text{ASP}$ and $[^{15}\text{N}]\text{ASP}$, respectively.

in the m/z domain as a superposition of two Gaussian distributions according to equation (5.5):

$$I(m/z) = \frac{A_1}{2\pi\sigma} \exp \frac{(x - \mu_1)^2}{2\sigma^2} + \frac{A_2}{2\pi\sigma} \exp \frac{(x - \mu_2)^2}{2\sigma^2} \quad (5.5)$$

where A_1 and A_2 denote the mixture coefficients for both Gaussians. If the mass peaks of individual isobars are Gaussian in shape, both can be fitted to the observed peak shape to derive the relative contributions of each species. In this context, the position of each peak corresponds to the exact mass of the considered isotope which can be easily calculated *a priori*. Since the RP_{FWHM} is constant in the considered small domain of the m/z axis, the individual peak widths of both isobars can be assumed to be identical. Therefore, the optimization problem needs to optimize three variables, i.e. the relative contribution of each species, A_1 and A_2 , and the peak width σ of both peaks.

14.4.2 Identification of unique mass transitions

Given the product ion spectrum of an amino acid, the question arises as to which fragment ion can be used for spectral deconvolution. When a ^{13}C - as well as ^{15}N -labeled sample is measured, the observed product ion spectrum for any given Q1 isolation window is a composite spectrum of all co-isolated ^{13}C - and ^{15}N -MI species. A fragment ion can only be used for spectral evaluation if to each of its labeling states always only one precursor MI contributes. If several MI species contribute to one of its labeling states, these cannot be disentangled. This problem is illustrated for the $[\text{M}+\text{H}]^+$ ion of ASP in Figure 5.6. In this figure the leftmost column designates a Q1 isolation window up to a mass shift of $k = 2$, followed by all possible precursor ion MI species that are co-isolated in this window for the $[\text{M}+\text{H}]^+$ ion of ASP. Subsequently, all labeling states for two product ion are shown.

It becomes apparent that to the $m[0,0]$ product ion MI of the m/z 43.0178 fragment in the isolation windows corresponding to $k = 1$ and $k = 2$ both precursor ion MI species contribute. This can be explained by the fact that this fragment does not retain the nitrogen atom of the precursor ion, thus each precursor MI with labeled nitrogen automatically contributes to the mass peak corresponding to a product ion MI with zero mass shift of nitrogen. A different picture emerges for the m/z 74.0237 fragment. For this fragment each precursor ion contributes to a distinct labeling state, although some cannot be resolved. In Q1 isolation window 135, for instance, singly-labeled ^{13}C -labeled ASP contributes to m/z 74.0237 and m/z 75.0270, whereas ^{15}N -ASP contributes to m/z 75.0207. The latter two peaks cannot be resolved in the TOF detector and need therefore to be deconvoluted.

More formally, the set of ^{13}C -labeling states in a fragment ion of given elemental composition and precursor ion of given elemental composition was derived in section 13.2.2 and is given in equation (5.1). This equation holds equally true for labeled nitrogen, in fact for any labeled element. Therefore, for nitrogen the following set describes the nitrogen labeling states in product ions:

$$TMI_N(\lambda^{P,N}, \kappa^{P,N}, \kappa^{F,N}) = \{\lambda^{F,N} \in \mathbb{N} \mid \lambda^{F,N} \geq \max([\lambda^{P,N} - (\kappa^{P,N} - \kappa^{F,N}), 0]) \wedge \lambda^{F,N} \leq \min([\kappa^{F,N}, \lambda^{P,N}])\} \quad (5.6)$$

14. QTOF MS TECHNOLOGY FOR SIMULTANEOUS ¹³C- AND ¹⁵N-ISOTOPE ENRICHMENT PROFILING

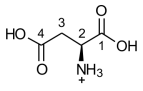
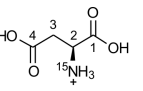
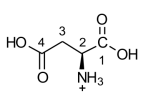
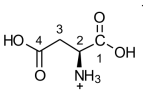
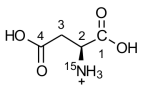
	Co-isolated Precursor Mass Isotopomers	Admissible labeling states of fragment 74.0237	Admissible labeling states of fragment 43.0178
Q1=134	 $C_4H_8NO_4^+$ 134.0448 M[0,0]	 $C_2H_4NO_2^+$ 74.0237 m[0,0]	 $C_2H_3O^+$ 43.0178 m[0,0]
Q1=135	 $C_4H_8^{15}NO_4^+$ 135.0418 M[0,1]	 $C_2H_4^{15}NO_2^+$ 75.0207 m[0,1]	 $C_2H_3O^+$ 43.0178 m[0,0]
	 $C_3^{13}CH_8NO_4^+$ 135.0481 M[1,0]	 $C^{13}H_4NO_2^+$ 74.0237 m[0,0]	 $C^{13}H_3O^+$ 43.0178 m[0,0]
Q1=136	 $2^{13}C$ $C_2^{13}C_2H_8NO_4^+$ 136.0515 M[2,0]	 $C_2H_4NO_2^+$ 74.0237 m[0,0]	 $C_2H_3O^+$ 43.0178 m[0,0]
		 $C^{13}H_4NO_2^+$ 75.0270 m[1,0]	 $C^{13}CH_3O^+$ 44.0212 m[1,0]
	 $C_3^{13}CH_8^{15}NO_4^+$ 136.0452 M[1,1]	 $^{13}C_2H_4NO_2^+$ 76.0304 m[2,0]	 $^{13}C_2H_3O^+$ 45.0246 m[2,0]
		 $C_2H_4^{15}NO_2^+$ 75.0207 m[0,1]	 $C_2H_3O^+$ 43.0178 m[0,0]

Figure 5.6: All possible labeling states for the precursor ion of $[M+H]^+$ of ASP up to a mass shift of $k = 2$. For two fragments of ASP all possible labeling states are listed arising from each possible precursor ion MI. In case several positions exist for a ¹³C-label, a bracket indicates the labeling state without specifying its position. Each structure is followed by its chemical formula and exact mass as well as the designator of the precursor ion MI.

Since the labeling in each element is independent from the one in other elements, each combination of ¹³C- and ¹⁵N-labeling states is admissible in any given fragment. Therefore, the number of possible fragment labeling states is given by:

$$TMI_{N-C} = \{(\lambda^{F,C}, \lambda^{F,N}) \in \mathbb{N} \times \mathbb{N} \mid \lambda^{F,C} \in TMI_C \wedge \lambda^{F,N} \in TMI_N\} \quad (5.7)$$

If the set of fragment labeling states is generated for the fragments m/z 74.0237 and m/z 43.017 according to equation (5.7) the labeling states of Figure 5.6 arise. Continuing this analysis for all

possible mass shifts of the $[M+H]^+$ ion of ASP reveals that to all admissible product ion labeling states just one precursor ion contributes. To identify all fragments, all precursor-product mass transitions can be computed given equation (5.7). First, all precursor ion MI species in the current Q1 window are computed. Then, for each precursor ion MI, all labeling states for each of its fragments are computed. The result is a list of precursor-product mass transitions, and a list of possible mass peaks (product ion mass isotopomers) in the spectrum of any Q1 window. From this list, it is easy to identify fragments to whose labeling states always only one precursor ion MI contributes. It appears that all fragment ions retaining the full set of nitrogen atoms and some part of the carbon skeleton fulfill the criterion that always only one precursor ion contributes to each of its labeling states.

14.5 Validation using defined ^{13}C - and ^{15}N -MI mixtures

To validate the approach of analytical separation using CID fragmentation, a series of defined isotopomer mixtures of $[4-^{13}\text{C}]\text{ASP}$ and $[^{15}\text{N}]\text{ASP}$ (at 100 μM total ASP concentration) were injected into the LC-ESI-QqTOF system running with the HPLC method of Appendix 9. ASP was measured in negative ionization mode and the product ion spectrum for the Q1 isolation window of 133 Da was recorded. The ^{13}C -enrichment was quantified in the m/z 115.0037 fragment of Figure 5.5. The relative intensity of the mass peaks at m/z 115.0037 and at m/z 116.0070 is shown in Table 5.5. It can be seen that the measured fractional abundance of each mass trace corresponds well with the expected intensity distribution and thus validates the applied approach.

To validate the proposed scheme for spectral deconvolution, the same isotopic series from above was injected into the MS and ionized in positive ionization mode. In this mode, the $[M+H]^+$ ion of ASP generates a fragment at m/z 116.0342 corresponding to a loss of water and retention of the amino group. The ^{15}N -labeled peak has an exact mass of m/z 117.0313 and its $[4-^{13}\text{C}]$ -labeled version an exact mass of m/z 117.0376. The mass spectrum of $Q1=135$ was monitored during the elution, of which the excerpt from m/z 117.02 to 117.05 is shown in Figure 5.7. As evident from this figure, the experimental mass peaks (black stems) clearly shift along the m/z axis from the exact mass $[^{15}\text{N}]\text{ASP}$ towards the exact mass of $[4-^{13}\text{C}]\text{ASP}$ for an increasing relative share of the latter in the isotopic mixture.

However, a simple superposition of Gaussian distributions as suggested by Meija and Caruso [123] does not describe the peak shape very well, especially at intermediate isotopic mixtures. The best fit solutions for the Gaussian mixture model are plotted as continuous blue lines in Figure 5.7. The experimental peak shape of the isotopic mixtures appears more 'peaked' compared to a mixture of two Gaussians. The actual peak shape depends on a multitude of instrumental factors. Initial kinetic energy distributions and instrument jitters all contribute to a dispersion in arrival time distribution of measured product ions. All these influences require the numerical simulation of a highly non-trivial problem to obtain final peak shapes [68]. Especially, the ion reflectron in the TOF detector can be expected to perform a non-linear transformation on the arrival time distribution. Therefore, different models were empirically evaluated for the quality of the resulting fit to the experimental mass peak. A superposition of two Cauchy-Lorentz distributions raised to the power of n according to equation

(5.8) yielded good agreement between experimental and fitted peak shapes. Values between 16-20 for n yielded overall best agreement between experimental and best-fit solutions (Figure 5.7). The relative contributions of the considered isotopic species to the mass peak in each sample determined by fitting equation (5.8) to the experimental mass peak are given in Table 5.5.

$$I(m/z) = \left[\frac{A_1}{\gamma\pi[1 + (\frac{m/z - m/z_{exact,15N}}{\gamma})^2]} + \frac{A_2}{\gamma\pi[1 + (\frac{m/z - m/z_{exact,13C}}{\gamma})^2]} \right]^n \quad (5.8)$$

It appears that the analytical separation based on CID fragmentation reflects the isotopic composition of the measured samples quite accurately. The discrepancy between estimated isotopic composition and the predicted fractional abundance lies in the range of 0.5 mol%. The spectral deconvolution approach, however, appears to yield only a rough estimate of the isotopic composition. This approach appears to suffer from inaccuracies in the range of 5 to 10 mol%. This is too great an error to use these data for ^{13}C -MFA.

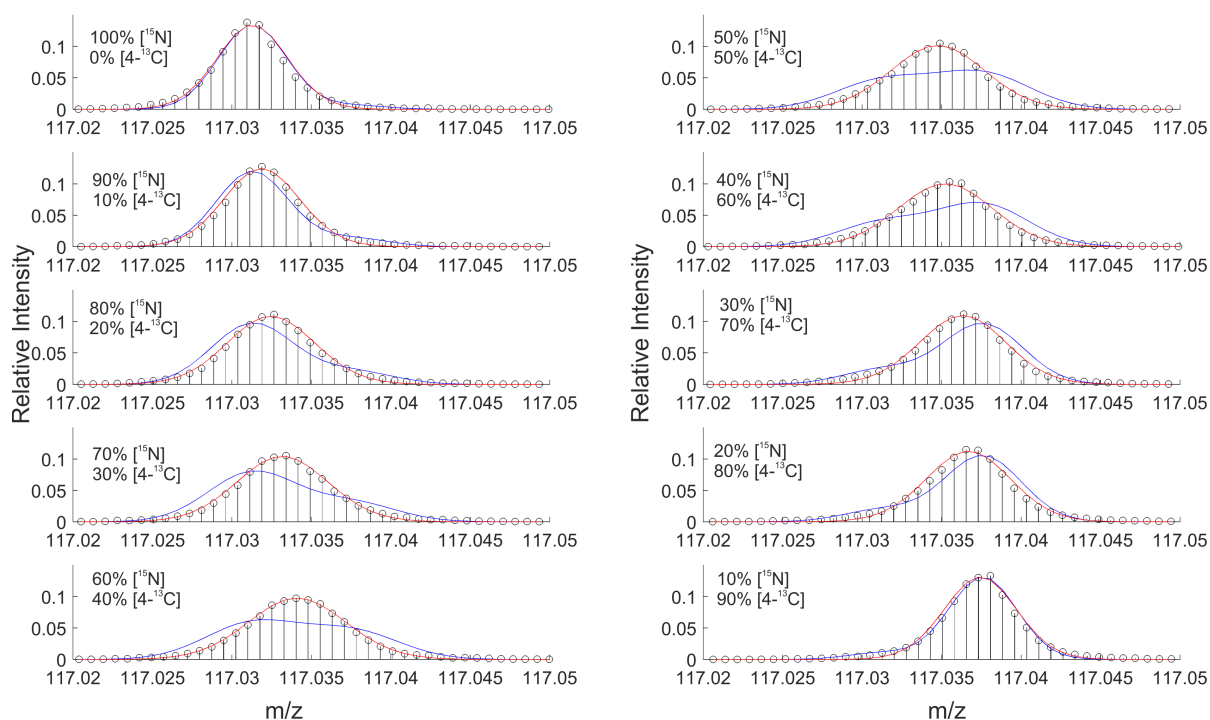


Figure 5.7: The experimental mass peak of the $[\text{M}-\text{H}_2\text{O}+\text{H}]^+$ fragment of the $[\text{M}+\text{H}]^+$ ion of ASP for different isotopic compositions of a 100 μM solution of ASP. Blue continuous line: Best fit to the experimental mass spectrum using a Gaussian mixture model according to equation (5.5). Red continuous line: Best fit of equation (5.8) to the experimental mass spectrum.

Table 5.5: Validation of the analytical separation of MI species by means of CID and the spectral deconvolution approach. Left column: Nominal composition of the isotopic mixture of $[4\text{-}^{13}\text{C}]\text{ASP}$ and $[^{15}\text{N}]\text{ASP}$. Middle columns: Relative intensities of the $m[0,0]$ tandem mass isotopomer with respect to the sum of the $m[0,0]$ and $m[1,0]$ tandem mass isotopomers in the TOF product ion spectrum of the Q1 isolation window of 133, corresponding to a singly-labeled $[\text{M-H}]^-$ ion of ASP. The standard deviation was determined from three replicate injections of the same sample. Experimental relative abundances and standard deviations thereof are given bold. The nominal isotopic composition corrected for natural isotopic abundance is given in italic in parenthesis following the experimental value. Rightmost columns: Minimal residual solutions for parameters of equation (5.8), fitted to the experimental mass peaks of Figure 5.7. The sum of parameters A_1 and A_2 was constrained to be 1.

Nominal isotopic composition	Approach based on CID Fragmentation		Approach based on spectral deconvolution	
	$[\text{M-NH}_3\text{-H}]^-$		$[\text{M-H}_2\text{O+H}]^+$	
	$\frac{m[0,0]}{(m[0,0] + m[1,0])}$	standard deviation	A_1	γ
100% ^{15}N 0% $4\text{-}^{13}\text{C}$	1 (<i>1</i>)	0	1	1.440e-02
90% ^{15}N 10% $4\text{-}^{13}\text{C}$	8.96e-01 (<i>8.99e-01</i>)	1.766e-03	8.83e-01	1.452e-02
80% ^{15}N 20% $4\text{-}^{13}\text{C}$	7.95e-01 (<i>7.99e-01</i>)	2.65e-03	7.88e-01	1.642e-02
70% ^{15}N 30% $4\text{-}^{13}\text{C}$	6.89e-01 (<i>6.98e-01</i>)	1.983e-03	6.79e-01	1.678e-02
60% ^{15}N 40% $4\text{-}^{13}\text{C}$	5.93e-01 (<i>5.98e-02</i>)	3.576e-03	5.48e-01	1.789e-02
50% ^{15}N 50% $4\text{-}^{13}\text{C}$	4.88e-01 (<i>4.98e-01</i>)	1.332e-03	4.56e-01	1.705e-02
40% ^{15}N 60% $4\text{-}^{13}\text{C}$	3.85e-01 (<i>3.98e-01</i>)	2.793e-03	3.89e-01	1.737e-02
30% ^{15}N 70% $4\text{-}^{13}\text{C}$	2.86e-01 (<i>2.98e-01</i>)	1.4034e-03	2.28e-01	1.619e-02
20% ^{15}N 80% $4\text{-}^{13}\text{C}$	1.88e-01 (<i>1.99e-01</i>)	4.801e-03	1.71e-01	1.596e-02
10% ^{15}N 90% $4\text{-}^{13}\text{C}$	9.39e-02 (<i>9.94e-02</i>)	5.328e-04	3.43e-01	1.415e-02
0% ^{15}N 100% $4\text{-}^{13}\text{C}$	0 (<i>4.19e-05</i>)	0	5.66e-05	1.404e-02

Conclusions And Outlook

This study provides a thorough investigation of the anaplerotic reactions in *C. glutamicum*, from an experimental as well as theoretical point of view. The theoretical analysis investigated the flux identifiability problem of the anaplerotic node using a combination of various structural FIA approaches differing in scope and accuracy. These approaches were combined to show in full generality that a comprehensive model of the central metabolism of *C. glutamicum* does not admit a unique solution for the fluxes at the anaplerotic node using stationary ^{13}C -MFA. This result holds globally, i.e. for all admissible substrates and even for the case of full measurement information. A reduction of the available information to measurable linear combinations of cumomer fractions has the same degree of information irrespective of the underlying network topology. The identifiable combinations of measurements and network topologies correspond to deletion mutants in *C. glutamicum*, but represent the native case in other prokaryotes. As such, the presented analysis is directly transferable to the anaplerotic node of other compartment-free organisms.

The results from the structural FIA inspired a list of anaplerotic deletion mutants that are structurally amenable to an estimation of intracellular fluxes by stationary ^{13}C -MFA. This set of mutant strains was complemented by other deletion mutants to provide a consistency check. The metabolism and adaptation of each mutant was characterized in a detailed manner by quantification of the extracellular specific rates, the intracellular metabolome of the central carbon metabolism by LC-QqQ MS and a large part of the proteome by SWATH acquisition using a LC-QqTOF MS platform. To harvest the wealth of measurements contained in the latter dataset, a regression approach suited for MS/MS proteome data was developed. This approach yielded precise and accurate fold changes for almost a third of all open reading frames of *C. glutamicum*. All datasets in conjunction consistently show that the carboxylation activities of PCx and PEPCx can be substituted for by the glyoxylate shunt under solely glycolytic conditions in *C. glutamicum*, under which condition the corresponding genes were thought to be repressed. It is inferred that this adaptation must be due to a changed interaction of a metabolite effector with the regulator protein RamA, brought about by a concentration change of the former. This metabolite pool, however, could not be identified from the set of metabolites from glycolysis, TCA and amino acids that was targeted in this study. Further research, especially based on untargeted metabolomics, will be needed to identify the metabolite regulator active under these conditions.

A subsequent ^{13}C -MFA study yielded inconclusive flux distributions that were inconsistent across the set of investigated strains. One possible error source may consist in strain-wise different biomass compositions, the quantification of which was considered to be beyond the scope of this study due to the complexity of the involved methods. Future work must include the establishment of protocols for routine quantification of biomass constituents in the scope of a ^{13}C -MFA. Apart from data refinement, additional efforts may be directed to the validation of the central carbon metabolism network of *C. glutamicum*. In particular, the absence of the activity of a particular label flow can be validated on

specially designed isotopomer mixtures in selected deletion mutants.

As one possible error source consists in the acquisition of the ^{13}C -label enrichment, the second part of the thesis was devoted to the development of a novel ^{13}C -profiling technique based on QqTOF MS technology which achieves an unprecedented positional resolution of ^{13}C -label within primary metabolites. It is shown that QqTOF MS technology represents the technological match to the demands of positionally resolved ^{13}C -labeling measurement, not only with respect to the speed of full spectral acquisition, but also with regard to the resolution of many interfering mass traces. To provide the basis for the positional assignment of ^{13}C -enrichment measurement, the product ion spectra of 76 precursor ions were studied with regard to their elemental composition as well as the positional origin of their carbon atoms. It was shown that nominal QqQ mass peaks can be composed of fragments of different elemental composition, compromising the measurement of the TMID in any of these fragments using QqQ MS technology with unit mass resolution. On the other hand, the signal intensity of some fragments with an uniquely defined elemental composition may draw from the contribution of two product ions with differing carbon origins, as shown by specifically synthesized ^{13}C -isotopic standards. These tracers also helped to expose that heuristic fragmentation rules to annotate CID product ion spectra with structures are often inept to predict the true fragment structure. In total, the elucidated carbon fate in this thesis represents the most accurate and comprehensive account yet on this matter. The inclusion of the ^{13}C -enrichment of a selection of the fragments whose carbon origin was identified or validated in this study improves flux confidence intervals, in particular for fluxes at the anaplerotic node.

In an exploratory last part of this thesis, the CID product ion spectra of amino acids were screened for dissociation events which allow to analytically separate ^{13}C - and ^{15}N -MI species without resolving them in the m/z domain. As a proof of concept, the separation of these isotopic species is shown for ASP. Although the concept is successfully applied to ASP, the $[\text{M}-\text{H}]^-$ and $[\text{M}+\text{H}]^+$ ions of just a few other amino acids undergo the necessary fragmentation event to enable ^{13}C - and ^{15}N -separation. Therefore, further research must be directed towards optimization of the parameters of the ion optics and fragmentation cell to enforce the necessary fragmentation.

List of Figures

1.1	Anaplerotic Node in <i>C. glutamicum</i>	2
1.2	Outline of this thesis	3
2.1	Overview of the central carbon metabolism of <i>C. glutamicum</i>	5
2.2	TCA operation under aerobic conditions when D-glucose is the carbon source	9
2.3	TCA operation under aerobic conditions when acetate is the carbon source	10
2.4	TCA operation under microaerobic conditions	12
2.5	Schematic representation of an ESI-QqQ MS device	15
2.6	Schematic representation of an ESI-QqTOF MS device	19
2.7	Exemplary CID reaction of peptides	24
2.8	QqTOF MS-based proteomics workflow employing SWATH acquisition	25
2.9	Workflow of a comprehensive ¹³ C-MFA study	29
2.10	Identifiability Concepts in ¹³ C-MFA	32
2.11	Relation between isotopomers of aspartate and tandem mass isotopomers of the <i>m/z</i> 88 fragment of the aspartate	37
4.1	Comprehensive network of the central carbon metabolism of <i>C. glutamicum</i>	65
4.2	Focused Network of the Anaplerotic Node in <i>C. glutamicum</i>	68
4.3	Global Identifiability of central metabolic reactionf for the wild-type and ΔPEPCx network	75
4.4	Global Identifiability of central metabolic reactions for the ΔPEPCx ME network	76
4.5	Adaptation of the Δ(PCx PEPCx) mutant on CGXII medium with glucose as sole carbon source	80
4.6	Representative fermentation for the Δ PEPCK and Δ(PCx PEPCx) strains	84
4.7	Estimated extracellular rates for anaplerotic deletion mutants	85
4.8	CO ₂ -formation rate and carbon equivalents of the specific growth and uptake rates for anaplerotic deletion mutants	87
4.9	pH-dependence of the forward and backward rate constants of the hydration of CO ₂ in see water at 30°C.	91
4.10	Simulated time course of CO _{2(aq)} and HCO ₃ ⁻ as well as the true and estimated CO ₂ -formation rates for three different scebarios of bicarbonate formation	94
4.11	Time course of the volume percentage of CO ₂ in the exhaust gas stream for four consid-ered cultivation regimes	95
4.12	Time course of the instantaneous carbon balance for <i>C. glutamicum</i> WT under different pH and aeration conditions	96
4.13	Intracellular concentrations of selected organic acids, amino acids and ACCOA in anaplerotic deletion mutants	100

4.14 Intracellular concentrations of selected organic and amino acids in anaplerotic deletion mutants	101
4.15 Specific extracellular concentrations of selected amino and organic acids for anaplerotic deletion mutants	103
4.16 Overview over MS/MS proteome data sets from SWATH MS acquisition	104
4.17 Score plot of the PCA on summed ¹⁵ N mass trace intensities	106
4.18 Score plot of the PCA on summed ¹⁵ N mass trace intensities	107
4.19 Loading plot for the second principle component of the PCA on summed ¹⁴ N-mass traces per protein	107
4.20 Validation of the Poisson regression approach on a isotopic mixture of known fold change	114
4.21 Fold changes of the enzymes PCx and ME across the set of anaplerotic deletion mutants	115
4.22 Volcano plot for the Δ(PCx PEPCx) strain showing higher abundant proteins in comparison to the WT	116
4.23 Fold changes of the acetate assimilation enzymes MS, ICL, ACK and PTA	118
4.24 Relative Share of significantly altered genes from each regulon of <i>C. glutamicum</i> in anaplerotic deletion mutants	120
4.25 ¹³ C-label enrichment in selected metabolite pools and multivariate ANOVA thereof . .	122
4.26 ¹³ C-label enrichment in metabolite pools pyruvate and malate	124
4.27 Estimated Flux Values for the WT, ΔPEPCx, ΔPCx and ΔME strains	127
4.28 Estimated Flux Values for the ΔPEPCk, Δ(PEPCk ME), Δ(PEPCx ME) and Δ(PCx ODx) strains	128
5.1 Notation for carbon labeling states used in this thesis	131
5.2 Exemplary accurate mass peaks collapsing onto one nominal mass peak	134
5.3 CID fragmentation of the [M-H] ⁻ ion of citrate in tabular form using structure names .	145
5.4 Statistical information gain from positionally resolved ¹³ C-labeling measurements . . .	150
5.5 Possible labeling states of the <i>m/z</i> 115.0037 product ion of ASP for ¹³ C- and ¹⁵ N-labeled aspartate for the three first quadrupole isolation windows	154
5.6 Possible labeling states of ¹³ C- and ¹⁵ N-labeled ASP for the three first Quadrupole isolation windows	159
5.7 Deconvolution of the singly-labeled <i>m/z</i> 116.0342 ion of ASP into isobaric contributions	161
1 Correlation between BV concentration and CDW concentration for <i>C. glutamicum</i> in CGXII medium without MOPS	172
2 Sum of Squared Residuals from the flux estimation in a <i>C. glutamicum</i> mutant library .	179
3 Exemplary acquisition of a amino acid standard mixture	181
4 L/H ratio along a dilution series of L-isoleucine and L-tyrosine measured on the LC-QqTOF system	182

LIST OF FIGURES

5	L/H ratio along a dilution series of L-glutamine and L-glutamate measured on the LC-QqTOF system	183
6	L/H ratio along a dilution series of L-valine and L-lysine measured on the LC-QqTOF system	184
7	L/H ratio along a dilution series of L-methionine and L-threonine measured on the LC-QqTOF system	185
8	L/H ratio along a dilution series of L-alanine and L-aspartate measured on the LC-QqTOF system	186
9	L/H ratio along a dilution series of L-leucine and L-phenylalanine measured on the LC-QqTOF system	187
10	L/H ratio along a dilution series of L-glycine, L-serine, and L-proline measured on the LC-QqTOF system	188

List of Tables

3.1	Consumable chemicals, solutions and non-standard materials.	46
3.2	Defined CGXII medium for bioreactor cultivations.	47
3.3	CGXII media for non-pH-regulated cultivations systems.	48
3.4	List of devices used in this work.	49
3.5	List of HPLC columns used in this work.	50
3.6	List of software used for the present work.	51
3.7	Gradient elution for amino acid quantification and ^{13}C -profiling.	56
3.8	HPLC gradients for the elution of sugar and nucleoside phosphates for quantification and ^{13}C -enrichment profiling.	57
3.9	HPLC method for SWATH Acquisition.	58
4.1	Identifiability of anaplerotic reactions in <i>C. glutamicum</i> under full information	69
4.2	Identifiability of anaplerotic reactions in <i>C. glutamicum</i> under partial information with either glucose as sole carbon substrate or during glucose and lactate co-metabolization	73
4.3	Genotype and reference for the used strains in this work	77
4.4	Identified insertions and deletions in the $\Delta(\text{PCx PEPCx})$ strain adapted to glucose as sole carbon source in comparison to <i>C. glutamicum</i> WT	81
4.5	Identified SNPs in the $\Delta(\text{PCx PEPCx})$ strain adapted to glucose as sole carbon source in comparison to <i>C. glutamicum</i> WT	82
5.1	Relative peak areas of individual mass traces for selectively ^{13}C -labeled 6PG, F6P, and FBP138	
5.2	Relative peak areas of selected mass traces for selectively ^{13}C -labeled OAA, MAL, αKG , GLU and ASP	141
5.3	Relative signal intensities of CID product ions of the M0, M1 and M2 mass isotopomers of MAL. A horizontal line indicates that the corresponding labeling state is not admissible, whereas '<LOD' (limit of detection) means that no mass peak could be integrated.	148
5.4	Relative intensities of product ions of amino acids corresponding to neutral loss combinations that leave the carbon backbone intact, but eliminate all nitrogen atoms. The intensity was quantified relative to the most intense product ion over a CE ramp of $\pm 5\text{-}50\text{ eV}$	155
5.5	Validation of two approaches to separate ^{13}C - and ^{15}N -MI species	162
1	Primers to verify gene deletions.	172
2	Estimated specific growth rates (μ), glucose consumption rates $\Pi_{\text{Glc,spec}}$ and CO_2 production rates ($\Pi_{\text{CO}_2,\text{spec}}$) for different deletion mutants	173
3	Parameters describing the CO_2 -dynamics in a batch process	175
4	SWATH Q1 windows monitored during elution on the TripleTOF 6600	176

LIST OF TABLES

5	Dummy variables coding scheme for the regression model to estimate L/H ratios. The colon sign denotes a nesting relationship.	177
6	Design Matrix using 'reference' coding. Again, the colon operator denotes a nesting relationship.	178
7	HPLC gradient used for ¹³ C-isotopomer profiling on the QqTOF MS	180

Appendices

1 Primers used to verify deletion mutant identity

Table 1: Primers to verify gene deletions.

deletion	Forward (sense) primer	Reverse (anti-sense) primer
$\Delta malE$ (cg3335)	CTT CCA GAC ACG GAA TCA GAG	GTG ATC CTT CCG AGC GTT CC
Δodx (cg1458)	ACC GGC ATC AAA TTG TGT C	TTG CCT TGA GCA CAAT GTC
Δppc (cg1787)	GTG AAC AGG CTC TCG ATG C	GGA ATA GAC TCG CTC GGC
Δpyc (cg0791)	GCA GAT GCC ATT TAC CCG	CGG TGA CAG ACT CAA CG
Δpck (cg3169)	GTT GAT GGA TCC CAG GCT GAG TG	GAA CTG GCT GTG AAC CTC TGC AG

2 Correlation between BV concentration and CDW concentration

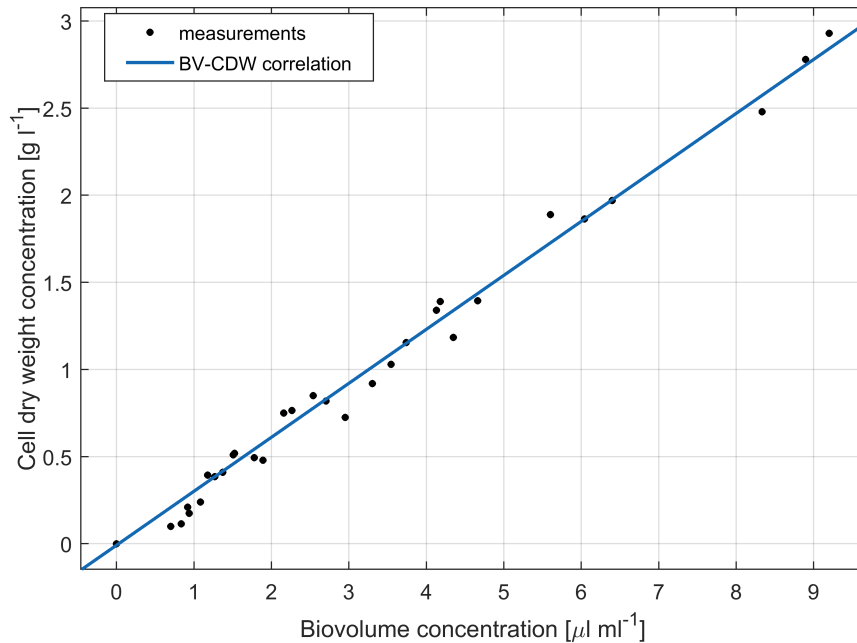


Figure 1: Correlation between BV and CDW concentration for *C. glutamicum* in CGXII medium without MOPS. The correlation depicts data points from the WT, Δ (PEPCx ME) and Δ (PCx ODx) strains grown on CGXII defined medium. The fitted coefficients for the function $f(x) = p1 \cdot x + p2$ are as follows (coefficient estimate followed by 95 % confidence interval): slope $p1 = 0.3098$ (0.3045, 0.3151) and intercept $p2 = -0.008692$ (-0.02189, 0.004509). The goodness of fit is SSE: 0.2256, R-square: 0.994, Adjusted R-square: 0.9939, RMSE: 0.05278.

3 Model-based estimates for specific rates in anaplerotic deletion mutants

Table 2: Estimated specific growth rates (μ), glucose consumption rates $\Pi_{\text{Glc,spec}}$ and CO_2 production rates ($\Pi_{\text{CO}_2,\text{spec}}$) for different deletion mutants. The units are provided in the nomenclature. Rates based on CDW bear the unit mmol (g h)^{-1} . The rates based on BV have the unit $\text{mmol ml}^{-1} \text{h}^{-1}$. The third and fourth columns contains the estimates of the specific growth rate by a simple exponential fit to the indicated biomass signal. The fifth and sixth columns contain the estimates obtained from the coupled inference of specific growth and uptake rate according to equations (3.6) and (3.7). Θ denotes the instantaneous carbon balance according to equation (4.5).

strain	Rep	μ (only BV)	μ (only CDW)	μ (BV)	μ (CDW)	$\Pi_{\text{Glc,spec}}$ (BV)	$\Pi_{\text{Glc,spec}}$ (CDW)	$\Pi_{\text{CO}_2,\text{spec}}$ (BV)	$\Pi_{\text{CO}_2,\text{spec}}$ (CDW)	Θ (CDW)
WT	140907/R7	0.45	0.45	0.45 [0.44, 0.46]	0.44 [0.42, 0.45]	1.47 [1.45, 1.50]	4.86 [4.68, 5.05]	2.20	7.16 $\pm$ 0.60	0.81 $\pm$ 0.077
	140714/R5	0.46	0.46	0.45 [0.44, 0.46]	0.45 [0.42, 0.48]	1.41 [1.36, 1.46]	4.61 [4.40, 4.85]	1.92	6.41 $\pm$ 0.67	0.84 $\pm$ 0.087
	141014/R6	0.46	0.46	0.47 [0.45, 0.48]	0.46 [0.42, 0.49]	1.53 [1.48, 1.58]	4.98 [4.77, 5.23]	2.23	7.24 $\pm$ 0.71	0.82 $\pm$ 0.089
$\Delta(\text{PCx ODCx})$	140714/R8	0.40	0.41	0.39 [0.38, 0.40]	0.40 [0.38, 0.42]	1.40 [1.38, 1.43]	4.54 [4.37, 4.70]	2.00	6.51 $\pm$ 0.96	0.78 $\pm$ 0.077
	140527/R8	0.39	0.39	0.38 [0.38, 0.38]	0.38 [0.37, 0.39]	1.31 [1.30, 1.32]	4.15 [4.04, 4.28]	1.87	6.07 $\pm$ 0.99	0.82 $\pm$ 0.074
	141208/R6	0.42	0.42	0.43 [0.42, 0.44]	0.42 [0.40, 0.44]	1.26 [1.22, 1.30]	4.44 [4.31, 4.60]	1.78	6.36 $\pm$ 0.89	0.83 $\pm$ 0.078
$\Delta(\text{PEPCx ME})$	140907/R5	0.46	0.47	0.46 [0.45, 0.47]	0.43 [0.42, 0.44]	1.46 [1.43, 1.49]	5.03 [4.83, 5.25]	2.14	6.96 $\pm$ 0.29	0.77 $\pm$ 0.072
	140714/R6	0.43	0.43	0.42 [0.41, 0.42]	0.42 [0.41, 0.43]	1.48 [1.45, 1.51]	4.77 [4.57, 4.96]	1.89	6.16 $\pm$ 0.83	0.77 $\pm$ 0.073
	140714/R7	0.43	0.43	0.43 [0.42, 0.45]	0.43 [0.41, 0.46]	1.40 [1.35, 1.45]	4.73 [4.48, 4.96]	2.23	7.25 $\pm$ 1.92	0.83 $\pm$ 0.088
ΔPEPCk	140907/R6	0.39	0.39	0.36 [0.36, 0.37]	0.37 [0.37, 0.38]	1.26 [1.25, 1.27]	4.03 [3.93, 4.14]	1.63	5.32 $\pm$ 0.51	0.80 $\pm$ 0.072
	140907/R8	0.39	0.39	0.38 [0.37, 0.38]	0.39 [0.37, 0.40]	1.27 [1.25, 1.29]	4.16 [4.02, 4.29]	1.75	5.69 $\pm$ 0.53	0.81 $\pm$ 0.077
ΔPEPCx	141014/R5	0.45	0.45	0.45 [0.44, 0.47]	0.46 [0.42, 0.49]	1.57 [1.53, 1.61]	5.08 [4.89, 5.32]	2.43	7.60 $\pm$ 0.92	0.81 $\pm$ 0.083
	141014/R8	0.44	0.44	0.43 [0.42, 0.44]	0.43 [0.41, 0.46]	1.48 [1.43, 1.53]	5.16 [4.95, 5.38]	2.36	7.68 $\pm$ 1.09	0.77 $\pm$ 0.076
$\Delta(\text{PEPCk ME})$	141014/R7	0.44	0.45	0.45 [0.43, 0.47]	0.45 [0.41, 0.48]	1.50 [1.45, 1.54]	4.94 [4.72, 5.18]	2.30	7.45 $\pm$ 0.67	0.82 $\pm$ 0.087
	141118/R5	0.45	0.46	0.47 [0.46, 0.48]	0.47 [0.45, 0.48]	1.48 [1.44, 1.52]	4.77 [4.58, 4.99]	2.26	7.37 $\pm$ 0.31	0.87 $\pm$ 0.083
	141118/R6	0.45	0.46	0.43 [0.43, 0.44]	0.45 [0.43, 0.47]	1.51 [1.49, 1.52]	4.85 [4.69, 5.02]	2.15	6.98 $\pm$ 0.83	0.81 $\pm$ 0.076
ΔPCx	141118/R3	0.41	0.41	0.38 [0.37, 0.39]	0.40 [0.38, 0.41]	1.44 [1.41, 1.46]	4.60 [4.45, 4.77]	2.24	7.29 $\pm$ 0.51	0.80 $\pm$ 0.072
	141118/R4	0.41	0.41	0.37 [0.36, 0.37]	0.40 [0.38, 0.41]	1.55 [1.53, 1.57]	4.70 [4.57, 4.84]	2.32	7.56 $\pm$ 0.76	0.79 $\pm$ 0.071
ΔME	141208/R5	0.47	0.48	0.48 [0.48, 0.49]	0.46 [0.45, 0.48]	1.50 [1.47, 1.52]	5.04 [4.98, 5.21]	2.02	7.36 $\pm$ 0.43	0.82 $\pm$ 0.075
	141208/R7	0.47	0.47	0.49 [0.48, 0.50]	0.46 [0.44, 0.48]	1.37 [1.36, 1.39]	4.72 [4.58, 4.87]	1.82	6.64 $\pm$ 1.00	0.85 $\pm$ 0.080
$\Delta(\text{PCx PEPCx})$	150428/R6	0.28	0.28	0.27 [0.27, 0.28]	0.27 [0.26, 0.28]	1.09 [1.06, 1.11]	3.62 [3.51, 3.77]	2.30	7.52 $\pm$ 0.68	0.81 $\pm$ 0.070
	150428/R7	0.28	0.28	0.26 [0.26, 0.28]	0.26 [0.25, 0.27]	1.20 [1.18, 1.23]	3.85 [3.69, 4.01]	2.23	7.27 $\pm$ 0.45	0.74 $\pm$ 0.067

4 Gaussian Error Propagation

$$\begin{aligned} \left[\Delta \left(\Pi_{\text{CO}_2, \text{spec}} + \frac{\omega_{\text{Carbon}} * 1000 * \mu}{M_{\text{Carbon}}} \right) \right]^2 &= \left| \frac{\omega_{\text{Carbon}} * 1000}{M_{\text{Carbon}}} \right|^2 (\Delta \mu)^2 \\ &+ \left| \frac{\mu * 1000}{M_{\text{Carbon}}} \right|^2 (\Delta \omega_{\text{Carbon}})^2 \\ &+ (\Delta \Pi_{\text{CO}_2, \text{spec}})^2 \end{aligned} \quad (1)$$

$$\begin{aligned} (\Delta \kappa)^2 &= \left| \frac{\omega_{\text{Carbon}} * 1000}{M_{\text{Carbon}} * (6 * \Pi_{\text{Glc, spec}})} \right|^2 (\Delta \mu)^2 \\ &+ \left| \frac{\mu * 1000}{M_{\text{Carbon}} * (6 * \Pi_{\text{Glc, spec}})} \right|^2 (\Delta \omega_{\text{Carbon}})^2 \\ &+ \left| \frac{1}{6 * \Pi_{\text{Glc, spec}}} \right|^2 (\Delta \Pi_{\text{CO}_2, \text{spec}})^2 \\ &+ \left| \frac{\left(\frac{\mu * \omega_{\text{Carbon}} * 1000}{M_{\text{Carbon}}} + \Pi_{\text{CO}_2, \text{spec}} \right)}{6 * (\Pi_{\text{Glc, spe}})^2} \right|^2 (\Delta \Pi_{\text{Glc, spe}})^2 \end{aligned} \quad (2)$$

5 Parameter values for the simulation study of CO₂ dynamics

Table 3: Parameters describing the CO₂-dynamics in a batch process.

parameter	description	unit	value
P_{ges}	total pressure	Pa	101325
T	temperature	K	302.15
R	universal gas constant	J mol ⁻¹ K ⁻¹	8.314
k^+	lumped forward rate constant	h ⁻¹	5.87e-02 · 3600 (pH=7)
k^-	lumped backward rate constant	h ⁻¹	5.32e-03 · 3600 (pH=7)
$k_L a$	overall volumetric mass transfer coefficient	h ⁻¹	8
K_{H,CO_2}	Henry's law volatility constant	l Pa mol ⁻¹	3.03e+06
$\Phi_{\text{CO}_2}^\alpha$	volume fraction of CO ₂ in inlet air	vol %	0.06
$\Phi_{\text{CO}_2}^\omega$	volume fraction of CO ₂ in outlet air	vol %	var.
$\Phi_{\text{O}_2}^\alpha$	volume fraction of O ₂ in inlet air	vol %	0.21
$\Phi_{\text{O}_2}^\omega$	volume fraction of O ₂ in outlet air	vol %	var.
F_{air}	aeration rate	m ³ h ⁻¹	12e-03
V_{liq}	reactor liquid volume	l	0.2
V_{air}	reactor gas phase volume	m ³	0.2e-03

6 Q1 Windows for SWATH Acquisition

Table 4: SWATH Q1 windows monitored during elution on the TripleTOF 6600. The SWATH windows were not scheduled.

SWATH Exp Index:	Start Mass (Da)	Stop Mass (Da)	CES
SWATH Exp 1:	199.50	222.10	5.000
SWATH Exp 2:	221.10	242.80	5.000
SWATH Exp 3:	241.80	264.40	5.000
SWATH Exp 4:	263.40	280.60	5.000
SWATH Exp 5:	279.60	295.00	5.000
SWATH Exp 6:	294.00	307.60	5.000
SWATH Exp 7:	306.60	319.30	5.000
SWATH Exp 8:	318.30	330.10	5.000
SWATH Exp 9:	329.10	340.00	5.000
SWATH Exp 10:	339.00	349.90	5.000
SWATH Exp 11:	348.90	358.90	5.000
SWATH Exp 12:	357.90	367.90	5.000
SWATH Exp 13:	366.90	376.00	5.000
SWATH Exp 14:	375.00	385.00	5.000
SWATH Exp 15:	384.00	393.10	5.000
SWATH Exp 16:	392.10	401.20	5.000
SWATH Exp 17:	400.20	410.20	5.000
SWATH Exp 18:	409.20	419.20	5.000
SWATH Exp 19:	418.20	428.20	5.000
SWATH Exp 20:	427.20	438.10	5.000
SWATH Exp 21:	437.10	448.90	5.000
SWATH Exp 22:	447.90	459.70	5.000
SWATH Exp 23:	458.70	472.30	5.000
SWATH Exp 24:	471.30	485.80	5.000
SWATH Exp 25:	484.80	500.20	5.000
SWATH Exp 26:	499.20	517.30	5.000
SWATH Exp 27:	516.30	537.10	5.000
SWATH Exp 28:	536.10	559.60	5.000
SWATH Exp 29:	558.60	584.80	5.000
SWATH Exp 30:	583.80	613.60	5.000
SWATH Exp 31:	612.60	645.10	5.000
SWATH Exp 32:	644.10	679.30	5.000
SWATH Exp 33:	678.30	718.90	5.000
SWATH Exp 34:	717.90	766.60	5.000
SWATH Exp 35:	765.60	826.00	5.000
SWATH Exp 36:	825.00	903.40	5.000
SWATH Exp 37:	902.40	1013.20	5.000
SWATH Exp 38:	1012.20	1171.60	5.000
SWATH Exp 39:	1170.60	1419.10	5.000
SWATH Exp 40:	1418.10	1998.70	5.000

7 Exemplary Design Matrix

A regression model with categorical variables can be specified by coding each level of a factor with a dummy variable which takes the value of 1 for all those observations corresponding to that level and 0 for all observations pertaining to other levels of that factor. The dummy variable coding for peak areas from two isotopes of two peptides for each of which two fragments are measured can be coded according to Table 5.

Table 5: Dummy variables coding scheme for the regression model to estimate L/H ratios. The colon sign denotes a nesting relationship.

Mean	Isotope		Peptide		Fragments:Pep1		Fragments:Pep2	
τ	isotope 1	isotope 2	Pep1	Pep2	Frag1: Pep1	Frag2:Pep1	Frag1:Pep2	Frag2:Pep2
1	1	0	1	0	1	0	0	0
1	1	0	1	0	0	1	0	0
1	1	0	0	1	0	0	1	0
1	1	0	0	1	0	0	0	1
1	0	1	1	0	1	0	0	0
1	0	1	1	0	0	1	0	0
1	0	1	0	1	0	0	1	0
1	0	1	0	1	0	0	0	1

Given the factor coding in Table 5, the coefficient τ has the interpretation of the overall mean. Coding Table 5 as matrix, we obtain the design matrix of equation (3). From its structure it becomes clear that the columns are linearly dependent. The binary variables for all levels of each factor added together always yield the variable coding the mean.

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \quad (3)$$

This can be circumvented by taking one level of each factor as reference, leaving just $n-1$ variables for each factor when n denotes the number of levels of that factor. In the following peptide 1 and isotope 1 are modeled as reference. With regard to the factor 'fragment' fragment 1 for each peptide is modeled as reference. Thereby, the effects of each fragment are always relative to one fragment within the same peptide group. Using this scheme, we obtain the dummy variable coding of Table 6.

From Table 6 it becomes apparent that we have $P * (F - 1)$ nested fragment variables, when the

Table 6: Design Matrix using 'reference' coding. Again, the colon operator denotes a nesting relationship.

Mean τ	Isotope isotope 2	Peptide Pep2	Fragments:Pep1 Frag2:Pep1	Fragments:Pep2 Frag2:Pep2
1	0	0	0	0
1	0	0	1	0
1	0	1	0	0
1	0	1	0	1
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0
1	1	1	0	1

data from P peptides and a constant number of fragments F per peptide are available. More generally, the regression problem has $\sum_p^P (F_p - 1)$ variables, when F_p denotes the total number of fragments for peptide p . Given the design matrix of Table 6, the regression equation for the running example becomes:

$$\mathbf{y} = \tau \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + I_2 \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \Pi_2 \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} + \Phi_{2(P_1)} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \Phi_{2(P_2)} \cdot \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad (4)$$

Using the reference coding scheme of equation (4), the meaning of τ has changed. It no longer represents the overall mean over data points. Evaluating equation (4) for a data point belonging to I_1 , P_1 , $F_1 : P_1$ yields:

$$y_{z, I_1, P_1, F_1(P_1)} = \tau \quad (5)$$

Thus, τ represents the mean of the reference group (coded with 0). Thereby, I_2 denotes the additive change of this group mean with respect to the I_1 group mean and P_2 the additive change with respect to the P_1 group mean and so on.

8 Sum of Squared Residuals from the flux estimation in a *C. glutamicum* mutant library

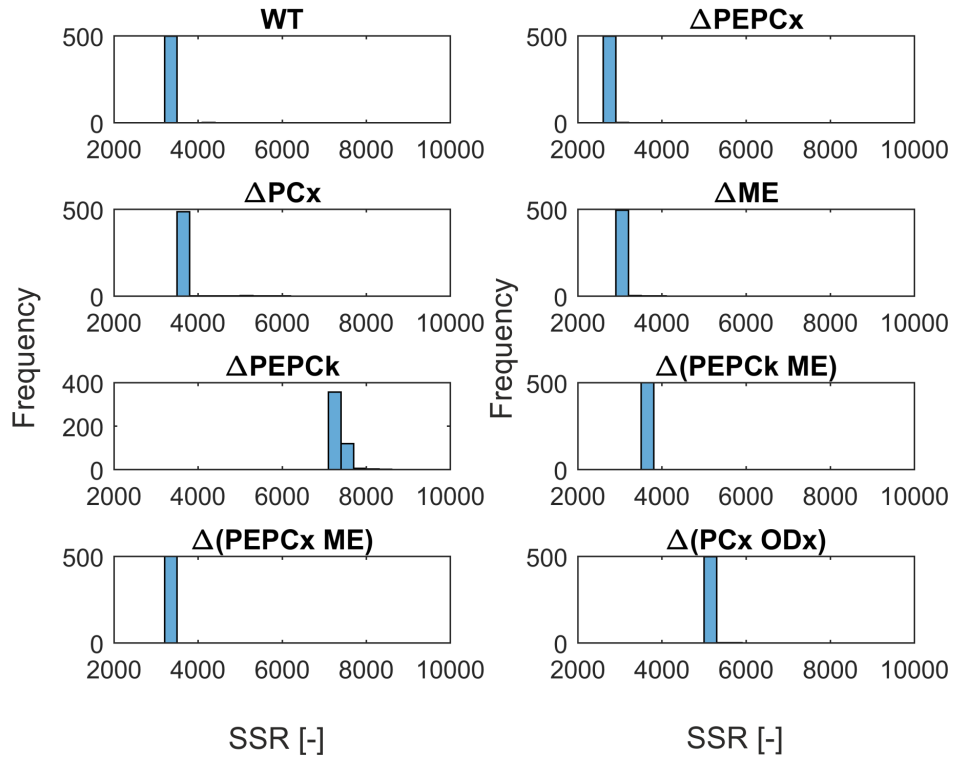


Figure 2: Sum of Squared Residuals from the flux estimation in a *C. glutamicum* mutant library. Each subplot shows the residuals of 500 solutions of the flux estimation in seven deletion mutants and the WT of *C. glutamicum*. Each solution represents the final optimizer output from 500 randomly initiated starting points.

9 HPLC method for a QqTOF acquisition method

Table 7: HPLC gradient used for ^{13}C -isotopomer profiling on the QqTOF MS. Contrary to the study of Teleki et al. [170], buffer A consisted of 100 % 10 mM ammonium acetate (pH 9.2) and buffer B of 100 % acetonitrile.

Time [min]	Flowrate [ml min ⁻¹]	% B
-15	0.2.	90
0	0.2	90
0.01	0.1	90
2	0.1	90
26	0.1	62
57	0.1	42
57.01	0.2	10
67	0.2	10
77	0.2	90
87	0.2	90

10 Exemplary acquisition of an amino acid standard mixture on the QqTOF device

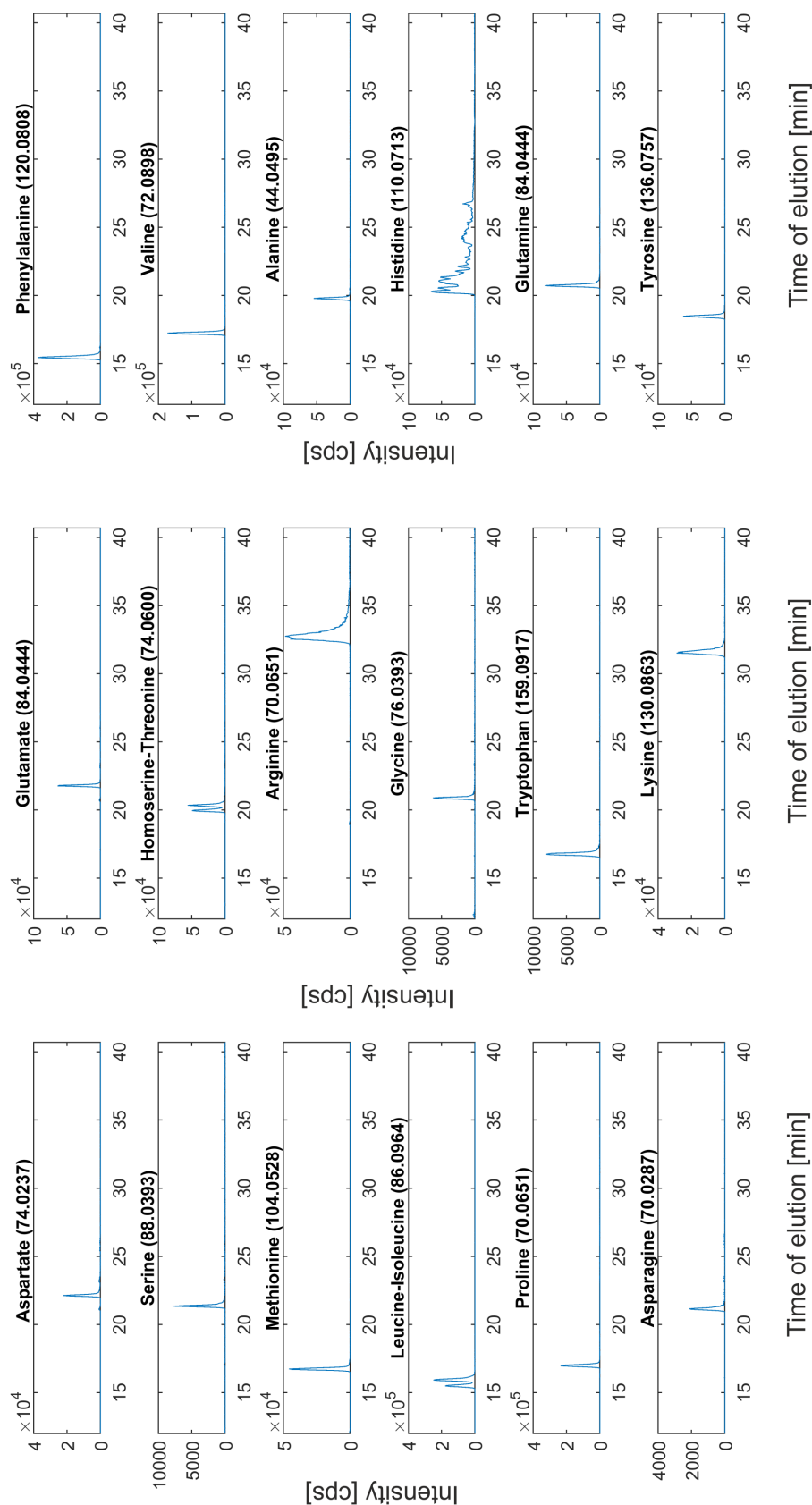


Figure 3: Exemplary acquisition of a 50 μ M amino acid standard mixture with the developed LC-QqTOF MS-based acquisition method. Shown are the total ion counts in the monitored Q1 windows. Injection volume was 5 μ L.

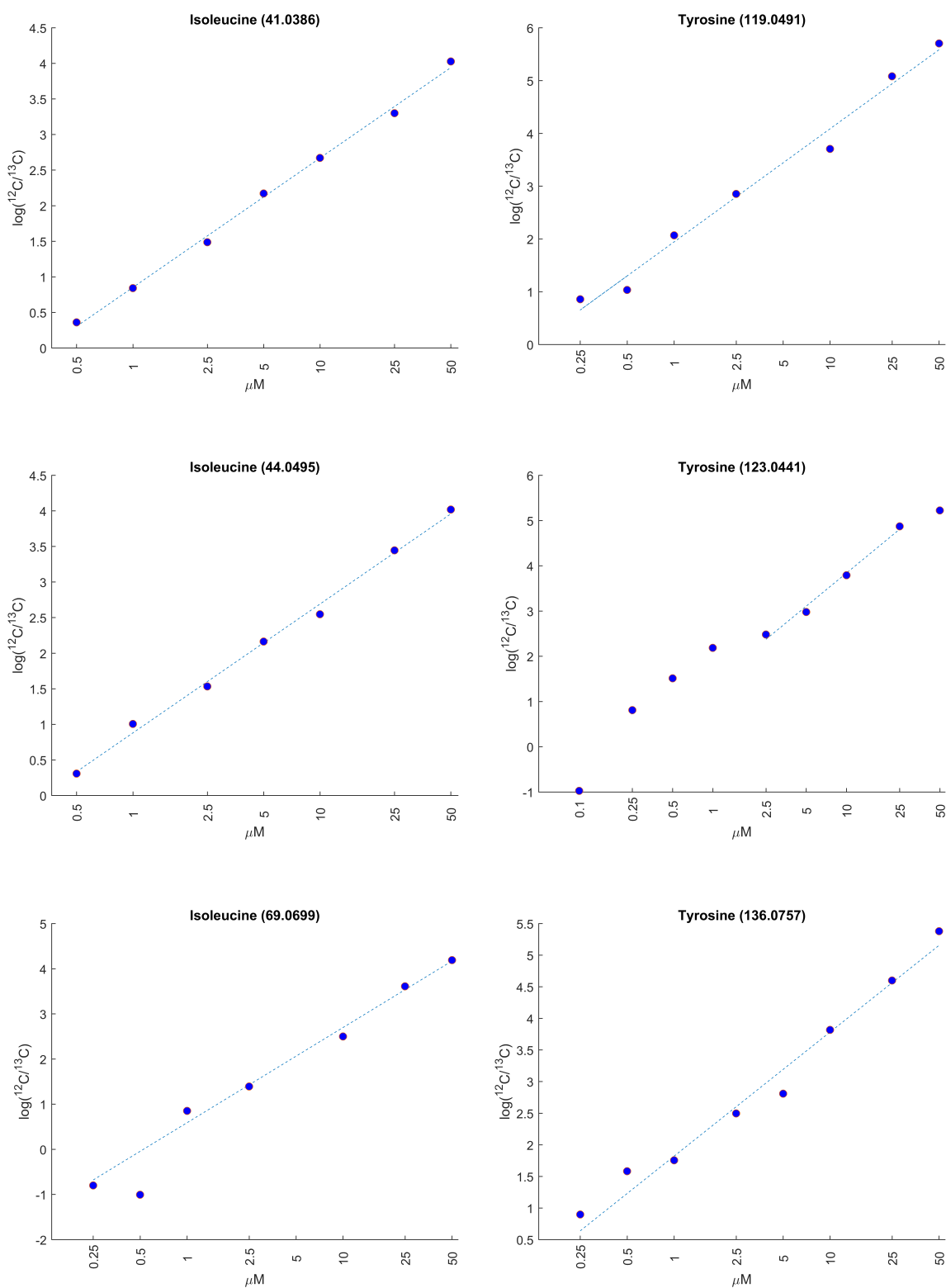


Figure 4: L/H ratio along a dilution series of L-isoleucine and L-tyrosine measured on the LC-QqTOF system.

10. EXEMPLARY ACQUISITION OF AN AMINO ACID STANDARD MIXTURE ON THE QQTOF DEVICE

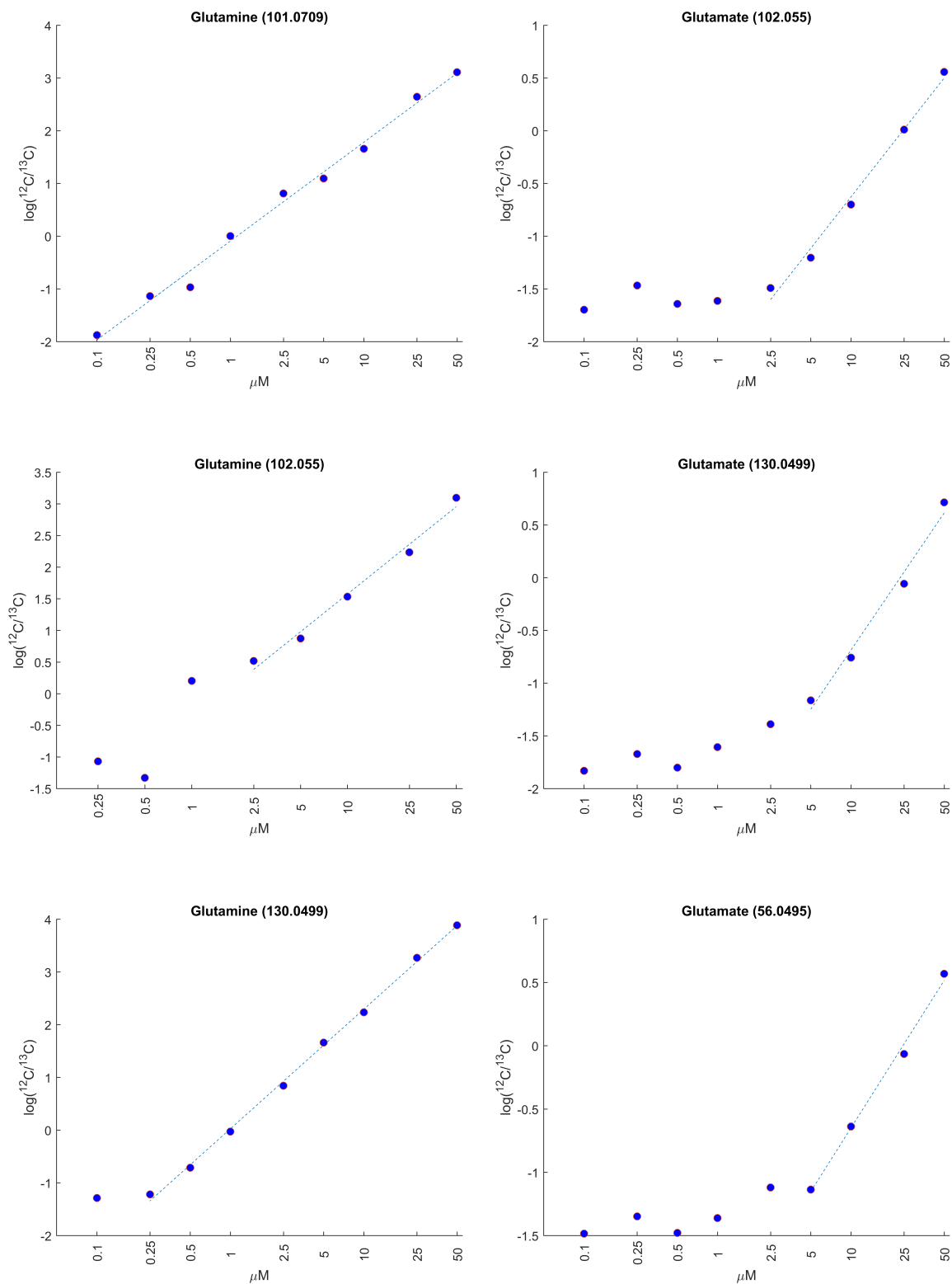


Figure 5: L/H ratio along a dilution series of L-glutamine and L-glutamate measured on the LC-QqTOF system.

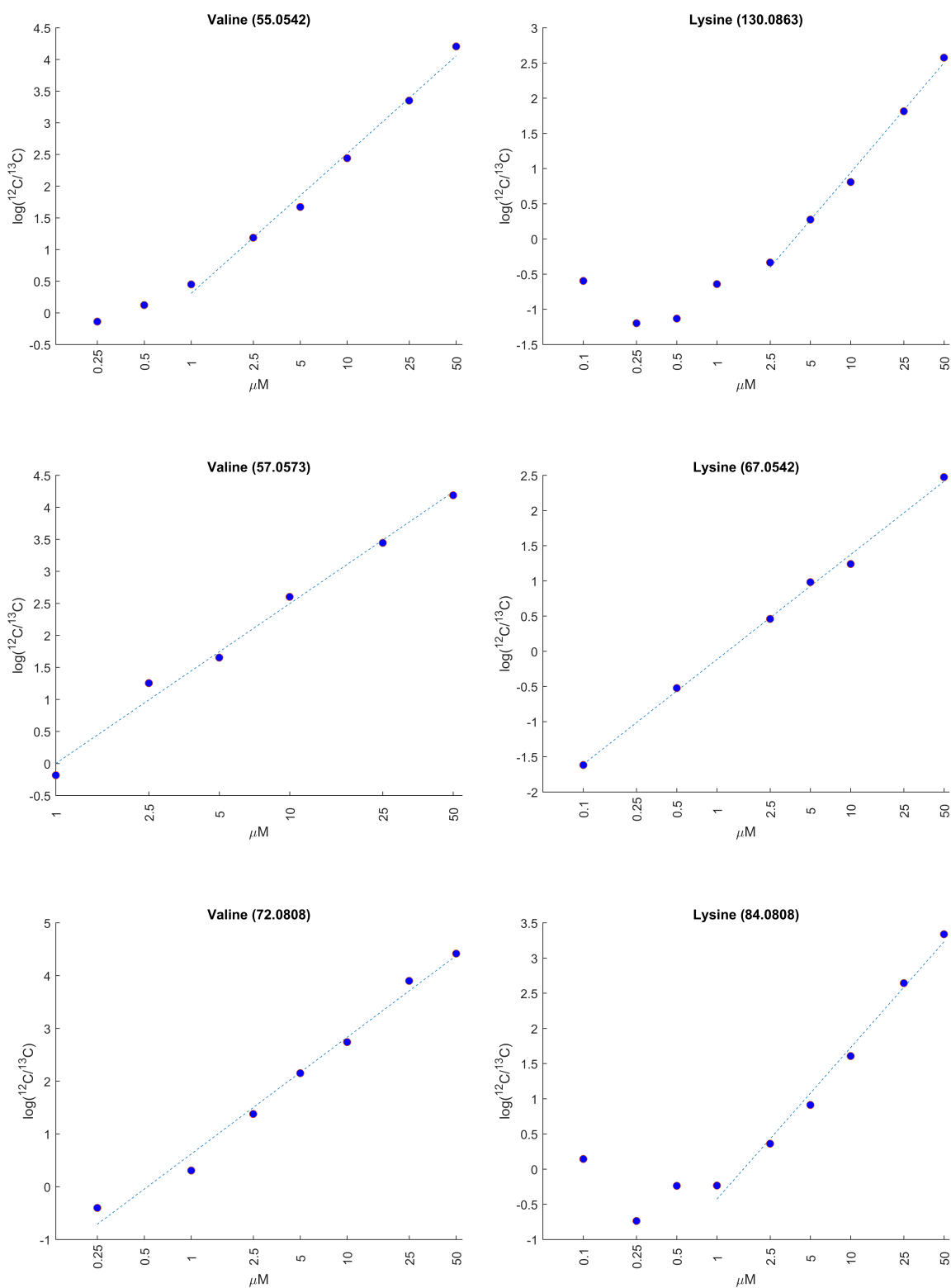


Figure 6: L/H ratio along a dilution series of L-valine and L-lysine measured on the LC-QqTOF system.

10. EXEMPLARY ACQUISITION OF AN AMINO ACID STANDARD MIXTURE ON THE QTOF DEVICE

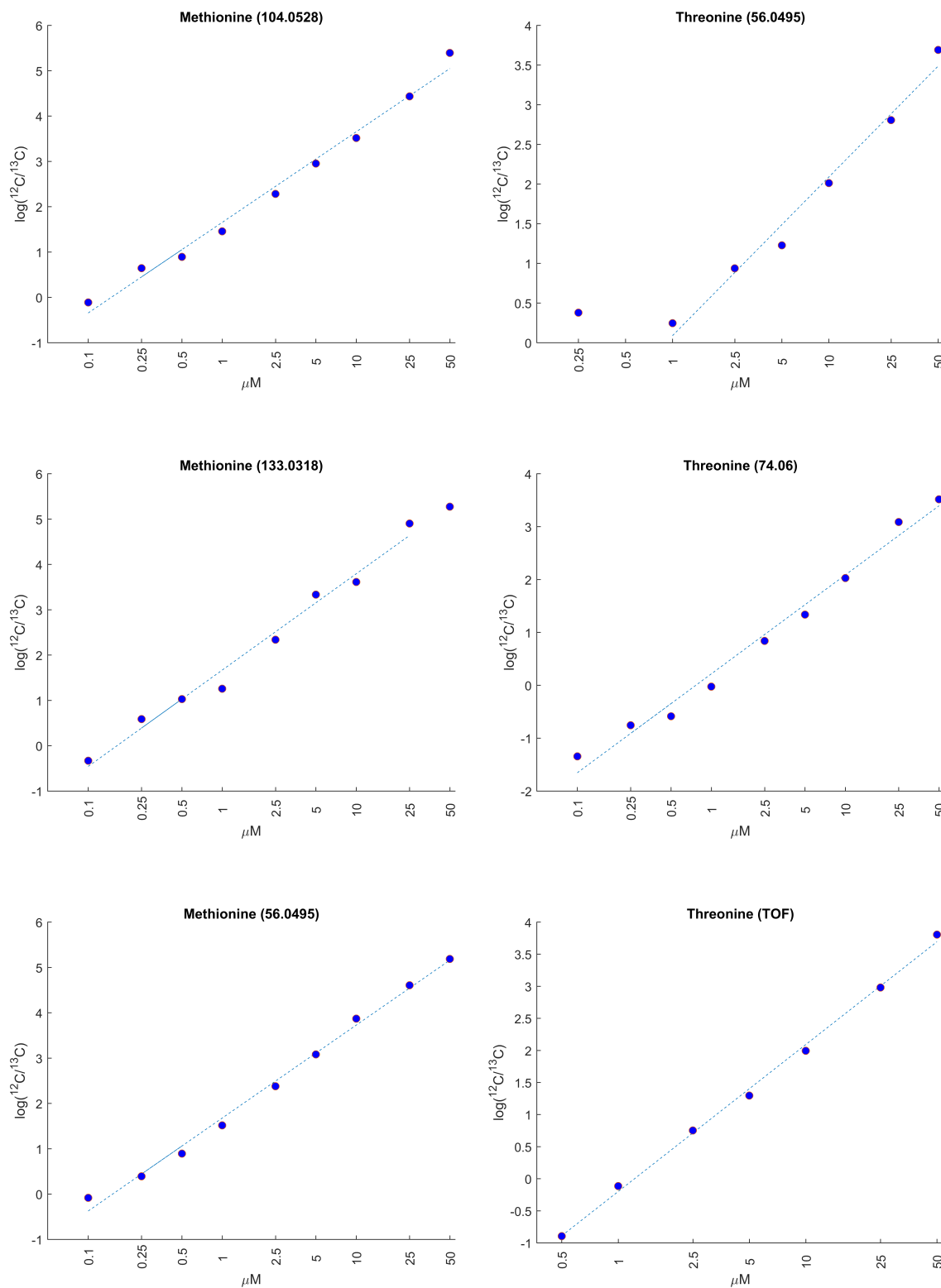


Figure 7: L/H ratio along a dilution series of L-methionine and L-threonine measured on the LC-QqTOF system.

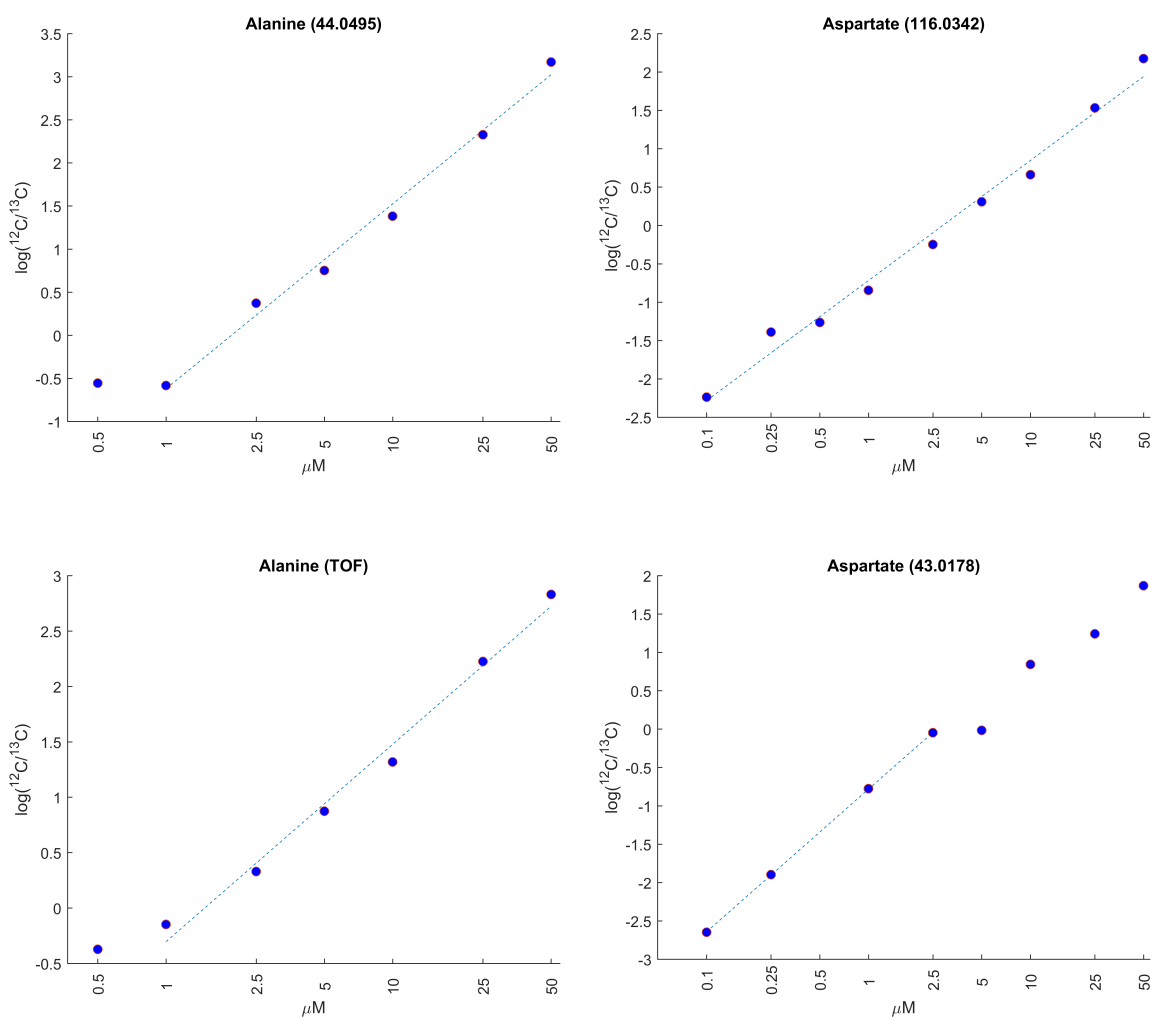


Figure 8: L/H ratio along a dilution series of L-alanine and L-aspartate measured on the LC-QqTOF system.

10. EXEMPLARY ACQUISITION OF AN AMINO ACID STANDARD MIXTURE ON THE QQTOF DEVICE

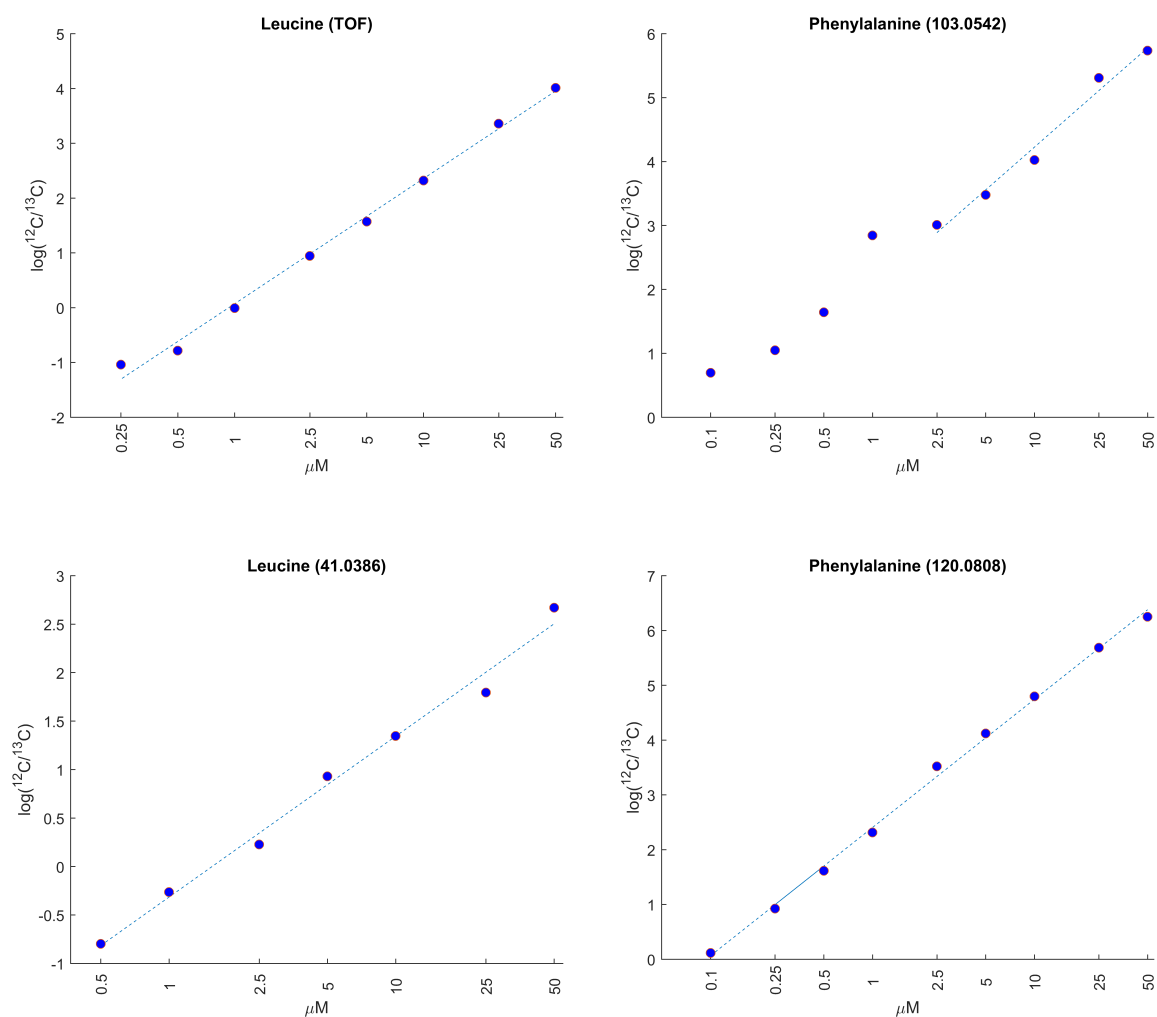


Figure 9: L/H ratio along a dilution series of L-leucine and L-phenylalanine measured on the LC-QqTOF system.

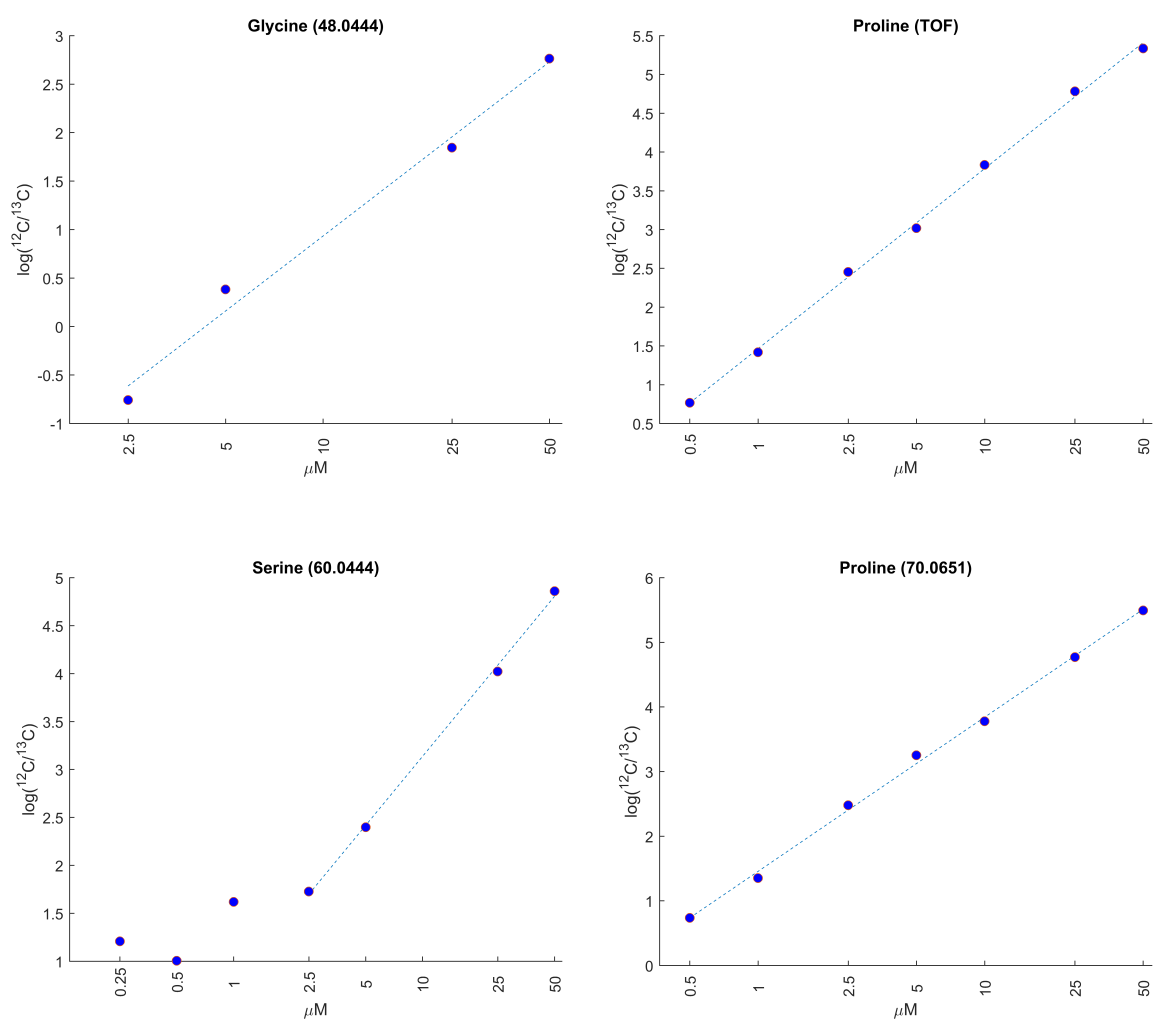


Figure 10: L/H ratio along a dilution series of L-glycine, L-serine, and L-proline measured on the LC-QqTOF system.

Bibliography

- [1] AB Sciex. *API 4000 LC/MS/MS Hardware Manual*. AB Sciex, 04 2010.
- [2] AB Sciex. *API 3200 Series of Instruments System User Guide*. AB Sciex, 04 2013.
- [3] AB Sciex. *TripleTOF 6600 System Planned Maintenance Procedure*. AB Sciex, 09 2016.
- [4] E. Ahadi and L. Konermann. Ejection of Solvated Ions from Electrosprayed Methanol/Water Nanodroplets Studied by Molecular Dynamics Simulations. *Journal of the American Chemical Society*, 133(24):9354–9363, 2011. doi: 10.1021/ja111492s. URL <http://dx.doi.org/10.1021/ja111492s>. PMID: 21591733.
- [5] F. Allen, A. Pon, M. Wilson, R. Greiner, and D. Wishart. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Research*, 2014. doi: 10.1093/nar/gku436. URL <http://nar.oxfordjournals.org/content/early/2014/06/03/nar.gku436.abstract>.
- [6] F. Allen, R. Greiner, and D. Wishart. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics*, 11(1):98–110, 2015. ISSN 1573-3882. doi: 10.1007/s11306-014-0676-4. URL <http://dx.doi.org/10.1007/s11306-014-0676-4>.
- [7] T. C. Alves, R. L. Pongratz, X. Zhao, O. Yarborough, S. Sereda, O. Shirihai, G. W. Cline, G. Mason, and R. G. Kibbey. Integrated, Step-Wise, Mass-Isotopomeric Flux Analysis of the TCA Cycle. *Cell Metabolism*, 22:936–947, 2015. doi: 10.1016/j.cmet.2015.08.021. URL <http://dx.doi.org/10.1016/j.cmet.2015.08.021>.
- [8] G. L. Andrews, B. L. Simons, J. B. Young, A. M. Hawkrigde, and D. C. Muddiman. Performance Characteristics of a New Hybrid Quadrupole Time-of-Flight Tandem Mass Spectrometer (TripleTOF 5600). *Analytical Chemistry*, 83(13):5442–5446, 2011. doi: 10.1021/ac200812d. URL <http://dx.doi.org/10.1021/ac200812d>.
- [9] M. R. Antoniewicz. ^{13}C metabolic flux analysis: optimal design of isotopic labeling experiments. *Current Opinion in Biotechnology*, 24(6):1116 – 1121, 2013. ISSN 0958-1669. doi: <http://dx.doi.org/10.1016/j.copbio.2013.02.003>. URL <http://www.sciencedirect.com/science/article/pii/S0958166913000177>.
- [10] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metabolic Engineering*, 9(1):68 – 86, 2007. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2006.09.001>. URL <http://www.sciencedirect.com/science/article/pii/S109671760600084X>.
- [11] J. Au, J. Choi, S. W. Jones, K. P. Venkataramanan, and M. R. Antoniewicz. Parallel labeling experiments validate *Clostridium acetobutylicum* metabolic network model for ^{13}C metabolic flux

- analysis. *Metabolic Engineering*, 26:23 – 33, 2014. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2014.08.002>. URL <http://www.sciencedirect.com/science/article/pii/S1096717614001050>.
- [12] M. Auchter, A. Cramer, A. Hüser, C. Rückert, D. Emer, P. Schwarz, A. Arndt, C. Lange, J. Kalinowski, V. F. Wendisch, and B. J. Eikmanns. RamA and RamB are global transcriptional regulators in *Corynebacterium glutamicum* and control genes for enzymes of the central metabolism. *Journal of Biotechnology*, 154(2–3):126 – 139, 2011. ISSN 0168-1656. doi: <http://dx.doi.org/10.1016/j.jbiotec.2010.07.001>. URL <http://www.sciencedirect.com/science/article/pii/S0168165610003007>.
- [13] S. U. Bajad, W. Lu, E. H. Kimball, J. Yuan, C. Peterson, and J. D. Rabinowitz. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *Journal of Chromatography A*, 1125(1):76 – 88, 2006. ISSN 0021-9673. doi: <http://dx.doi.org/10.1016/j.chroma.2006.05.019>. URL <http://www.sciencedirect.com/science/article/pii/S0021967306009939>.
- [14] S. Banerjee and S. Mazumdar. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *International Journal of Analytical Chemistry*, page 40 pages, 2012. doi: 10.1155/2012/282574. URL <https://www.hindawi.com/journals/ijac/2012/282574/>.
- [15] R. Bansal-Mutalik and H. Nikaido. Quantitative lipid composition of cell envelopes of *Corynebacterium glutamicum* elucidated through reverse micelle extraction. *Proceedings of the National Academy of Sciences*, 108(37):15360–15365, 2011. doi: 10.1073/pnas.1112572108. URL <http://www.pnas.org/content/108/37/15360.abstract>.
- [16] T. Bartek, B. Blombach, S. Lang, B. J. Eikmanns, W. Wiechert, M. Oldiges, K. Nöh, and S. Noack. Comparative ^{13}C metabolic flux analysis of pyruvate dehydrogenase complex-deficient, L-valine-producing *Corynebacterium glutamicum*. *Applied and environmental microbiology*, 77(18):6644–52, Sept. 2011. ISSN 1098-5336. doi: 10.1128/AEM.00575-11. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3187166&tool=pmcentrez&rendertype=abstract>.
- [17] J. Becker and C. Wittmann. Bio-based production of chemicals, materials and fuels – *Corynebacterium glutamicum* as versatile cell factory. *Current Opinion in Biotechnology*, 23(4):631 – 640, 2012. ISSN 0958-1669. doi: <http://dx.doi.org/10.1016/j.copbio.2011.11.012>. URL <http://www.sciencedirect.com/science/article/pii/S0958166911007233>.
- [18] J. Becker and C. Wittmann. GC-MS-Based ^{13}C Metabolic Flux Analysis. In J. O. Krömer, L. K. Nielsen, and L. M. Blank, editors, *Metabolic Flux Analysis*, volume 1191 of *Methods in Molecu-*

- lar Biology*, pages 165–174. Springer New York, 2014. ISBN 978-1-4939-1169-1. doi: 10.1007/978-1-4939-1170-7_10. URL http://dx.doi.org/10.1007/978-1-4939-1170-7_10.
- [19] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995. ISSN 00359246. URL <http://www.jstor.org/stable/2346101>.
- [20] D. J. Beste, K. Noh, S. Niedenfuhr, T. A. Mendum, N. D. Hawkins, J. L. Ward, M. H. Beale, W. Wiechert, and J. McFadden. ^{13}C -Flux Spectral Analysis of Host-Pathogen Metabolism Reveals a Mixed Diet for Intracellular *Mycobacterium tuberculosis*. *Chemistry and Biology*, 20(8): 1012 – 1021, 2013. ISSN 1074-5521. doi: <http://dx.doi.org/10.1016/j.chembiol.2013.06.012>. URL <http://www.sciencedirect.com/science/article/pii/S1074552113002457>.
- [21] M. Blein-Nicolas and M. Zivy. Thousand and one ways to quantify and compare protein abundances in label-free bottom-up proteomics. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1864(8):883 – 895, 2016. ISSN 1570-9639. doi: <https://doi.org/10.1016/j.bbapap.2016.02.019>. URL <http://www.sciencedirect.com/science/article/pii/S1570963916300322>.
- [22] B. Blombach, J. Buchholz, T. Busche, J. Kalinowski, and R. Takors. Impact of different CO_2/HCO_3 levels on metabolism and regulation in *Corynebacterium glutamicum*. *Journal of Biotechnology*, 168(4):331–340, 2013. ISSN 0168-1656. doi: <http://dx.doi.org/10.1016/j.jbiotec.2013.10.005>. URL <http://www.sciencedirect.com/science/article/pii/S016816561300429X>.
- [23] R. Boisseau, B. Charrier, S. Massou, J.-C. Portais, S. Akoka, and P. Giraudeau. Fast Spatially Encoded 3D NMR Strategies for ^{13}C -Based Metabolic Flux Analysis. *Analytical Chemistry*, 85(20): 9751–9757, 2013. doi: 10.1021/ac402155w. URL <http://dx.doi.org/10.1021/ac402155w>.
- [24] J. Buchholz, M. Graf, B. Blombach, and R. Takors. Improving the carbon balance of fermentations by total carbon analyses. *Biochemical Engineering Journal*, 90(0):162 – 169, 2014. ISSN 1369-703X. doi: <http://dx.doi.org/10.1016/j.bej.2014.06.007>. URL <http://www.sciencedirect.com/science/article/pii/S1369703X14001594>.
- [25] T. Cajka and O. Fiehn. Toward Merging Untargeted and Targeted Methods in Mass Spectrometry-Based Metabolomics and Lipidomics. *Analytical Chemistry*, 88(1):524–545, 2016. doi: 10.1021/acs.analchem.5b04491. URL <http://dx.doi.org/10.1021/acs.analchem.5b04491>.
- [26] T. Cajka and O. Fiehn. Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Analytical Chemistry*, 88(1):524–545, 2016. doi: 10.1021/acs.analchem.5b04491. URL <http://dx.doi.org/10.1021/acs.analchem.5b04491>. PMID: 26637011.

- [27] O. Chahrour, D. Cobice, and J. Malone. Stable isotope labelling methods in mass spectrometry-based quantitative proteomics. *Journal of Pharmaceutical and Biomedical Analysis*, 113:2 – 20, 2015. ISSN 0731-7085. doi: <http://dx.doi.org/10.1016/j.jpba.2015.04.013>. URL <http://www.sciencedirect.com/science/article/pii/S0731708515002460>.
- [28] J. Choi and M. R. Antoniewicz. Tandem mass spectrometry: A novel approach for metabolic flux analysis. *Metabolic Engineering*, 13(2):225 – 233, 2011. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2010.11.006>. URL <http://www.sciencedirect.com/science/article/pii/S1096717610001047>.
- [29] J. Choi, M. T. Grossbach, and M. R. Antoniewicz. Measuring Complete Isotopomer Distribution of Aspartate Using Gas Chromatography/Tandem Mass Spectrometry. *Analytical Chemistry*, 84(10):4628–4632, 2012. doi: 10.1021/ac300611n. URL <http://dx.doi.org/10.1021/ac300611n>.
- [30] D. B. Chu, C. Troyer, T. Mairinger, K. Ortmayr, S. Neubauer, G. Koellensperger, and S. Hann. Isotopologue analysis of sugar phosphates in yeast cell extracts by gas chromatography chemical ionization time-of-flight mass spectrometry. *Analytical and Bioanalytical Chemistry*, 407(10): 2865–2875, 2015. ISSN 1618-2650. doi: 10.1007/s00216-015-8521-9. URL <http://dx.doi.org/10.1007/s00216-015-8521-9>.
- [31] T. Clough, M. Key, I. Ott, S. Ragg, G. Schadow, and O. Vitek. Protein Quantification in Label-Free LC-MS Experiments. *Journal of Proteome Research*, 8(11):5275–5284, 2009. doi: 10.1021/pr900610q. URL <http://dx.doi.org/10.1021/pr900610q>.
- [32] T. Clough, S. Thaminy, S. Ragg, R. Aebersold, and O. Vitek. Statistical protein quantification and significance analysis in label-free LC-MS experiments with complex designs. *BMC Bioinformatics*, 13(16):1–17, 2012. ISSN 1471-2105. doi: 10.1186/1471-2105-13-S16-S6. URL <http://dx.doi.org/10.1186/1471-2105-13-S16-S6>.
- [33] M. Coccagn-Bousquet, A. Guyonvarch, and N. D. Lindley. Growth Rate-Dependent Modulation of Carbon Flux through Central Metabolism and the Kinetic Consequences for Glucose-Limited Chemostat Cultures of *Corynebacterium glutamicum*. *Applied and Environmental Microbiology*, 62(2):429–36, 1996. URL <http://aem.asm.org/content/62/2/429.abstract>.
- [34] L. T. Cordova, J. Lu, R. M. Cipolla, N. R. Sandoval, C. P. Long, and M. R. Antoniewicz. Co-utilization of glucose and xylose by evolved *Thermus thermophilus* LC113 strain elucidated by ¹³C metabolic flux analysis and whole genome sequencing. *Metabolic Engineering*, 37: 63 – 71, 2016. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2016.05.001>. URL <http://www.sciencedirect.com/science/article/pii/S1096717616300131>.
- [35] A. Cramer and B. J. Eikmanns. RamA, the Transcriptional Regulator of Acetate Metabolism in *Corynebacterium glutamicum*, Is Subject to Negative Autoregulation. *Journal of Molecular*

- Microbiology and Biotechnology*, 12:51–59, 2007. doi: 10.1159/000096459. URL <https://www.karger.com/Article/Abstract/96459>.
- [36] A. Cramer, M. Auchter, J. Frunzke, M. Bott, and B. J. Eikmanns. RamB, the Transcriptional Regulator of Acetate Metabolism in *Corynebacterium glutamicum*, Is Subject to Regulation by RamA and RamB. *Journal of Bacteriology*, 189(3):1145–1149, 2007. doi: 10.1128/JB.01061-06. URL <http://jlb.asm.org/content/189/3/1145.abstract>.
- [37] S. B. Crown and M. R. Antoniewicz. Selection of tracers for ^{13}C -Metabolic Flux Analysis using Elementary Metabolite Units (EMU) basis vector methodology. *Metabolic Engineering*, 14(2): 150 – 161, 2012. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2011.12.005>. URL <http://www.sciencedirect.com/science/article/pii/S1096717611001273>.
- [38] S. B. Crown, C. P. Long, and M. R. Antoniewicz. Integrated ^{13}C -metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metabolic Engineering*, 28:151 – 158, 2015. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2015.01.001>. URL <http://www.sciencedirect.com/science/article/pii/S1096717615000038>.
- [39] T. Dalman, T. Dörnemann, E. Juhnke, M. Weitzel, W. Wiechert, K. Nöh, and B. Freisleben. Cloud MapReduce for Monte Carlo bootstrap applied to Metabolic Flux Analysis. *Future Generation Computer Systems*, 29(2):582 – 590, 2013. ISSN 0167-739X. doi: <http://dx.doi.org/10.1016/j.future.2011.10.007>. URL <http://www.sciencedirect.com/science/article/pii/S0167739X11001920>.
- [40] E. de Hoffmann. Tandem mass spectrometry: A primer. *Journal of Mass Spectrometry*, 31(2):129–137, 1996. ISSN 1096-9888. doi: 10.1002/(SICI)1096-9888(199602)31:2<129::AID-JMS305>3.0.CO;2-T. URL [http://dx.doi.org/10.1002/\(SICI\)1096-9888\(199602\)31:2<129::AID-JMS305>3.0.CO;2-T](http://dx.doi.org/10.1002/(SICI)1096-9888(199602)31:2<129::AID-JMS305>3.0.CO;2-T).
- [41] L. Di Donato, C. Des Rosiers, J. A. Montgomery, F. David, M. Garneau, and H. Brunengraber. Rates of gluconeogenesis and citric acid cycle in perfused livers, assessed from the mass spectrometric assay of the ^{13}C labeling pattern of glutamate. *Journal of Biological Chemistry*, 268(6):4170–80, 1993. URL <http://www.jbc.org/content/268/6/4170.abstract>.
- [42] S. Di Palma, R. Raijmakers, A. J. R. Heck, and S. Mohammed. Evaluation of the Deuterium Isotope Effect in Zwitterionic Hydrophilic Interaction Liquid Chromatography Separations for Implementation in a Quantitative Proteomic Approach. *Analytical Chemistry*, 83(21):8352–8356, 2011. doi: 10.1021/ac2018074. URL <https://doi.org/10.1021/ac2018074>.
- [43] N. N. Dookeran, T. Yalcin, and A. G. Harrison. Fragmentation Reactions of Protonated α -Amino Acids. *Journal of Mass Spectrometry*, 31(5):500–508, 1996. doi: 10.1002/(SICI)1096-9888(199605)31:5<500::AID-JMS327>3.0.CO;2-Q. URL [http://dx.doi.org/10.1002/\(SICI\)1096-9888\(199605\)31:5<500::AID-JMS327>3.0.CO;2-Q](http://dx.doi.org/10.1002/(SICI)1096-9888(199605)31:5<500::AID-JMS327>3.0.CO;2-Q).

- [44] P. Droste, S. Miebach, S. Niedenführ, W. Wiechert, and K. Nöh. Visualizing multi-omics data in metabolic networks with the software omix—a case study. *Biosystems*, 105(2):154 – 161, 2011. ISSN 0303-2647. doi: <http://dx.doi.org/10.1016/j.biosystems.2011.04.003>. URL <http://www.sciencedirect.com/science/article/pii/S0303264711000761>.
- [45] M. Eckersley, J. H. Bowie, and R. N. Hayes. Collision-induced dissociations of deprotonated α -amino acids. The occurrence of specific proton transfers preceding fragmentation. *International Journal of Mass Spectrometry and Ion Processes*, 93(2):199 – 213, 1989. ISSN 0168-1176. doi: [http://dx.doi.org/10.1016/0168-1176\(89\)80097-7](http://dx.doi.org/10.1016/0168-1176(89)80097-7). URL <http://www.sciencedirect.com/science/article/pii/0168117689800977>.
- [46] L. Eggeling and M. Bott. *Handbook of Corynebacterium glutamicum*. CRC Press, Boca Raton, 2005.
- [47] Eichinger, Peter C. H. and Hayes, Roger N. and Bowie, John H. Anionic rearrangement in the gas phase. The collision-induced loss of carbon monoxide from deprotonated pyruvates and hydroxyacetates. *Journal of the Chemical Society, Perkin Transactions 2*, pages 1815–1819, 1990. doi: 10.1039/P29900001815. URL <http://dx.doi.org/10.1039/P29900001815>.
- [48] R. E. Ekman, J. E. Silberring, Westman-Brinkmalm, A. M. (Editor), A. Kraj, D. M. Desiderio, and N. M. Nibbering. *Mass Spectrometry. Instrumentation, Interpretation, and Applications*. Wiley Series on Mass Spectrometry. John Wiley and Sons Ltd, 2009.
- [49] H. El Aribi, G. Orlova, A. C. Hopkinson, and K. W. M. Siu. Gas-Phase Fragmentation Reactions of Protonated Aromatic Amino Acids: Concomitant and Consecutive Neutral Eliminations and Radical Cation Formations. *The Journal of Physical Chemistry A*, 108(17):3844–3853, 2004. doi: 10.1021/jp0374915. URL <http://dx.doi.org/10.1021/jp0374915>.
- [50] M. Esler, D. Griffith, S. Wilson, and L. Steele. Precision Trace Gas Analysis by FT-IR Spectroscopy. 2. The $^{13}\text{C}/^{12}\text{C}$ Isotope Ratio of CO_2 . *Analytical Chemistry*, 72(1):216–221, 2000. doi: 10.1021/ac990563x. URL <http://dx.doi.org/10.1021/ac990563x>.
- [51] G. Fedorova, T. Randak, R. H. Lindberg, and R. Grabic. Comparison of the quantitative performance of a Q-Exactive high-resolution mass spectrometer with that of a triple quadrupole tandem mass spectrometer for the analysis of illicit drugs in wastewater. *Rapid Communications in Mass Spectrometry*, 27(15):1751–1762, 2013. ISSN 1097-0231. doi: 10.1002/rcm.6628. URL <http://dx.doi.org/10.1002/rcm.6628>.
- [52] J. Fenn, M. Mann, C. Meng, S. Wong, and C. Whitehouse. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, 1989. ISSN 0036-8075. doi: 10.1126/science.2675315. URL <http://science.sciencemag.org/content/246/4926/64>.

- [53] A. Flamholz, E. Noor, A. Bar-Even, and R. Milo. eQuilibrator—the biochemical thermodynamics calculator. *Nucleic Acids Research*, 40 (Database issue):D770–D775, 2012. doi: <http://doi.org/10.1093/nar/gkr874>. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3245061/>.
- [54] L. V. Foster and T. A. Davis. Algorithm 933: Reliable Calculation of Numerical Rank, Null Space Bases, Pseudoinverse Solutions, and Basic Solutions Using suitesparseQR. *ACM Transactions on Mathematical Software*, 40(1):7:1–7:23, Oct. 2013. ISSN 0098-3500. doi: 10.1145/2513109.2513116. URL <http://doi.acm.org/10.1145/2513109.2513116>.
- [55] J. Fredens, K. Engholm-Keller, A. Giessing, D. Pultz, M. Rossel Larsen, P. Hojrup, J. Møller-Jensen, and N. J. Faergeman. Quantitative proteomics by amino acid labeling in *C. elegans*. *Nature Methods*, 8:845–847, 2011. doi: 10.1038/nmeth.1675. URL <http://www.nature.com/nmeth/journal/v8/n10/full/nmeth.1675.html>.
- [56] H.-P. Garnir, Y. Baudinet-Robinet, and P.-D. Dumont. Error in the determination of gaussian peak centroids. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 28(1):146 – 153, 1987. ISSN 0168-583X. doi: [http://dx.doi.org/10.1016/0168-583X\(87\)90051-6](http://dx.doi.org/10.1016/0168-583X(87)90051-6). URL <http://www.sciencedirect.com/science/article/pii/0168583X87900516>.
- [57] R. Gerstmeir, V. F. Wendisch, S. Schnicke, H. Ruan, M. Farwick, D. Reinscheid, and B. J. Eikmanns. Acetate metabolism and its regulation in *Corynebacterium glutamicum*. *Journal of Biotechnology*, 104(1–3):99 – 122, 2003. ISSN 0168-1656. doi: [http://dx.doi.org/10.1016/S0168-1656\(03\)00167-6](http://dx.doi.org/10.1016/S0168-1656(03)00167-6). URL <http://www.sciencedirect.com/science/article/pii/S0168165603001676>. A New Era in *Corynebacterium Glutamicum* Biotechnology.
- [58] K. Gevaert, F. Impens, B. Ghesquière, P. Van Damme, A. Lambrechts, and J. Vandekerckhove. Stable isotopic labeling in proteomics. *PROTEOMICS*, 8(23-24):4873–4885, 2008. ISSN 1615-9861. doi: 10.1002/pmic.200800421. URL <http://dx.doi.org/10.1002/pmic.200800421>.
- [59] B. Gibbons and J. Edsall. Rate of Hydration of Carbon Dioxide and Dehydration of Carbonic Acid at 25°C. *Journal of Biological Chemistry*, 238(10):3502–3507, 1963. URL <http://www.jbc.org/content/238/10/3502.short>.
- [60] H. G. Gika, G. A. Theodoridis, U. Vrhovsek, and F. Mattivi. Quantitative profiling of polar primary metabolites using hydrophilic interaction ultrahigh performance liquid chromatography–tandem mass spectrometry. *Journal of Chromatography A*, 1259:121 – 127, 2012. ISSN 0021-9673. doi: <http://dx.doi.org/10.1016/j.chroma.2012.02.010>. URL <http://www.sciencedirect.com/science/article/pii/S0021967312002531>.
- [61] L. C. Gillet, P. Navarro, S. Tate, H. Röst, N. Selevsek, L. Reiter, R. Bonner, and R. Aebersold. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A

- New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics*, 11(6), 2012. doi: 10.1074/mcp.O111.016717. URL <http://www.mcponline.org/content/11/6/0111.016717.abstract>.
- [62] G. L. Glish and D. J. Burinsky. Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *Journal of the American Society for Mass Spectrometry*, 19(2):161 – 172, 2008. ISSN 1044-0305. doi: <http://dx.doi.org/10.1016/j.jasms.2007.11.013>. URL <http://www.sciencedirect.com/science/article/pii/S1044030507009981>.
- [63] P. Gourdon, M.-F. Baucher, N. D. Lindley, and A. Guyonvarch. Cloning of the Malic Enzyme Gene from *Corynebacterium glutamicum* and Role of the Enzyme in Lactate Metabolism. *Applied and Environmental Microbiology*, 66(7):2981–2987, 2000. doi: 10.1128/AEM.66.7.2981-2987.2000. URL <http://aem.asm.org/content/66/7/2981.abstract>.
- [64] L. E. Greene, J. S. Grossert, and R. L. White. Correlations of ion structure with multiple fragmentation pathways arising from collision-induced dissociations of selected α -hydroxycarboxylic acid anions. *Journal of Mass Spectrometry*, 48(3):312–320, 2013. ISSN 1096-9888. doi: 10.1002/jms.3158. URL <http://dx.doi.org/10.1002/jms.3158>.
- [65] J. S. Grossert, P. D. Fancy, and R. L. White. Fragmentation pathways of negative ions produced by electrospray ionization of acyclic dicarboxylic acids and derivatives. *Canadian Journal of Chemistry*, 83(11):1878–1890, 2005. doi: 10.1139/v05-214. URL <http://dx.doi.org/10.1139/v05-214>.
- [66] J. S. Grossert, M. C. Cook, and R. L. White. The influence of structural features on facile McLafferty-type, even-electron rearrangements in tandem mass spectra of carboxylate anions. *Rapid Communications in Mass Spectrometry*, 20(10):1511–1516, 2006. ISSN 1097-0231. doi: 10.1002/rcm.2506. URL <http://dx.doi.org/10.1002/rcm.2506>.
- [67] M. Guilhaus. Special feature: Tutorial. Principles and instrumentation in time-of-flight mass spectrometry. Physical and instrumental concepts. *Journal of Mass Spectrometry*, 30(11):1519–1532, 1995. ISSN 1096-9888. doi: 10.1002/jms.1190301102. URL <http://dx.doi.org/10.1002/jms.1190301102>.
- [68] M. Guilhaus, V. Mlynski, and D. Selby. Perfect Timing: Time-of-flight Mass Spectrometry. *Rapid Communications in Mass Spectrometry*, 11(9):951–962, 1997. ISSN 1097-0231. doi: 10.1002/(SICI)1097-0231(19970615)11:9<951::AID-RCM785>3.0.CO;2-H. URL [http://dx.doi.org/10.1002/\(SICI\)1097-0231\(19970615\)11:9<951::AID-RCM785>3.0.CO;2-H](http://dx.doi.org/10.1002/(SICI)1097-0231(19970615)11:9<951::AID-RCM785>3.0.CO;2-H).
- [69] N. Gunnarsson, U. H. Mortensen, M. Sosio, and J. Nielsen. Identification of the Enter–Doudoroff pathway in an antibiotic-producing actinomycete species. *Molecular Microbiology*, 52(3):895–902, 2004. ISSN 1365-2958. doi: 10.1111/j.1365-2958.2004.04028.x. URL <http://dx.doi.org/10.1111/j.1365-2958.2004.04028.x>.

- [70] A. G. Harrison and T. Yalcin. Proton mobility in protonated amino acids and peptides. *International Journal of Mass Spectrometry and Ion Processes*, 165:339 – 347, 1997. ISSN 0168-1176. doi: [http://dx.doi.org/10.1016/S0168-1176\(97\)00173-0](http://dx.doi.org/10.1016/S0168-1176(97)00173-0). URL <http://www.sciencedirect.com/science/article/pii/S0168117697001730>.
- [71] M. Heinonen, A. Rantanen, T. Mielikäinen, J. Kokkonen, J. Kiuru, R. A. Ketola, and J. Rousu. FiD: a software for ab initio structural identification of product ions from tandem mass spectrometric data. *Rapid Communications in Mass Spectrometry*, 22(19):3043–3052, 2008. ISSN 1097-0231. doi: 10.1002/rcm.3701. URL <http://dx.doi.org/10.1002/rcm.3701>.
- [72] S. M. Herbrich, R. N. Cole, K. P. West, K. Schulze, J. D. Yager, J. D. Groopman, P. Christian, L. Wu, R. N. O’Meally, D. H. May, M. W. McIntosh, and I. Ruczinski. Statistical Inference from Multiple iTRAQ Experiments without Using Common Reference Standards. *Journal of Proteome Research*, 12(2):594–604, 2013. doi: 10.1021/pr300624g. URL <http://dx.doi.org/10.1021/pr300624g>.
- [73] S. Heux, J. Poinot, S. Massou, S. Sokol, and J.-C. Portais. A novel platform for automated high-throughput fluxome profiling of metabolic variants. *Metabolic Engineering*, 25:8 – 19, 2014. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2014.06.001>. URL <http://www.sciencedirect.com/science/article/pii/S1096717614000743>.
- [74] E. G. Hill, J. H. Schwacke, S. Comte-Walters, E. H. Slate, A. L. Oberg, J. E. Eckel-Passow, T. M. Therneau, and K. L. Schey. A Statistical Model for iTRAQ Data Analysis. *Journal of Proteome Research*, 7(8):3091–3101, 2008. doi: 10.1021/pr070520u. URL <http://dx.doi.org/10.1021/pr070520u>.
- [75] C. Ho and J. Sturtevant. The Kinetics of the Hydration of Carbon Dioxide at 25°C. *The Journal of Biological Chemistry*, 238(10):3499–3501, 1963. URL <http://www.jbc.org/content/238/10/3499.short>.
- [76] R. J. Holewinski, S. J. Parker, A. D. Matlock, and J. E. Venkatraman, Vidyaand Van Eyk. *Methods for SWATH™: Data Independent Acquisition on TripleTOF Mass Spectrometers*, pages 265–279. Springer New York, New York, NY, 2016. ISBN 978-1-4939-3524-6. doi: 10.1007/978-1-4939-3524-6_16. URL http://dx.doi.org/10.1007/978-1-4939-3524-6_16.
- [77] D. V. Hoof, M. W. H. Pinkse, D. W.-V. Oostwaard, C. L. Mummery, A. J. R. Heck, and J. Krijgsveld. An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nature Methods*, 4:677 – 678, 2007. doi: 10.1038/nmeth0907-677. URL <http://www.nature.com/nmeth/journal/v4/n9/full/nmeth0907-677.html>.
- [78] H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai,

- H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka. Massbank: a public repository for sharing mass spectral data for life sciences. *Journal of Mass Spectrometry*, 45(7):703–714, 2010. ISSN 1096-9888. doi: 10.1002/jms.1777. URL <http://dx.doi.org/10.1002/jms.1777>.
- [79] M. Hörl, J. Schnidder, U. Sauer, and N. Zamboni. Non-stationary ^{13}C -metabolic flux ratio analysis. *Biotechnology and Bioengineering*, 110(12):3164–3176, 2013. ISSN 1097-0290. doi: 10.1002/bit.25004. URL <http://dx.doi.org/10.1002/bit.25004>.
- [80] F. Hufsky, K. Scheubert, and S. Böcker. Computational mass spectrometry for small-molecule fragmentation. *TrAC Trends in Analytical Chemistry*, 53(0):41 – 48, 2014. doi: <http://dx.doi.org/10.1016/j.trac.2013.09.008>. URL <http://www.sciencedirect.com/science/article/pii/S016599361300232X>.
- [81] N. Isermann and W. Wiechert. Metabolic isotopomer labeling systems. Part II: structural flux identifiability analysis. *Mathematical Biosciences*, 183(2):175 – 214, 2003. ISSN 0025-5564. doi: [http://dx.doi.org/10.1016/S0025-5564\(02\)00222-5](http://dx.doi.org/10.1016/S0025-5564(02)00222-5). URL <http://www.sciencedirect.com/science/article/pii/S0025556402002225>.
- [82] L. J. Jazmin and J. D. Young. *Isotopically Nonstationary ^{13}C Metabolic Flux Analysis*, pages 367–390. Humana Press, Totowa, NJ, 2013. ISBN 978-1-62703-299-5. doi: 10.1007/978-1-62703-299-5_18. URL http://dx.doi.org/10.1007/978-1-62703-299-5_18.
- [83] F. H. Jeffrey, J. Roach, C. J. Storey, A. Sherry, and C. R. Malloy. ^{13}C Isotopomer Analysis of Glutamate by Tandem Mass Spectrometry. *Analytical Biochemistry*, 300(2):192 – 205, 2002. ISSN 0003-2697. doi: <http://dx.doi.org/10.1006/abio.2001.5457>. URL <http://www.sciencedirect.com/science/article/pii/S0003269701954570>.
- [84] K. S. Johnson. Carbon dioxide hydration and dehydration kinetics in seawater. *Limnology and Oceanography*, 27(5):849–855, 1982. ISSN 1939-5590. doi: 10.4319/lo.1982.27.5.0849. URL <http://dx.doi.org/10.4319/lo.1982.27.5.0849>.
- [85] E. Jolkver, D. Emer, S. Ballan, R. Krämer, B. J. Eikmanns, and K. Marin. Identification and Characterization of a Bacterial Transport System for the Uptake of Pyruvate, Propionate, and Acetate in *Corynebacterium glutamicum*. *Journal of Bacteriology*, 191(3):940–948, 2009. doi: 10.1128/JB.01155-08. URL <http://jlb.asm.org/content/191/3/940.abstract>.
- [86] I. Jorge, P. Navarro, P. Martínez-Acedo, E. Núñez, H. Serrano, A. Alfranca, J. M. Redondo, and J. Vázquez. Statistical Model to Analyze Quantitative Proteomics Data Obtained by $^{18}\text{O}/^{16}\text{O}$ Labeling and Linear Ion Trap Mass Spectrometry: Application to the Study of Vascular Endothelial Growth Factor-induced Angiogenesis in Endothelial Cells. *Molecular and Cellular Proteomics*,

- 8(5):1130–1149, 2009. doi: 10.1074/mcp.M800260-MCP200. URL <http://www.mcponline.org/content/8/5/1130.abstract>.
- [87] C. Jungreuthmayer, S. Neubauer, T. Mairinger, J. Zanghellini, and S. Hann. ICT : isotope correction toolbox. *Bioinformatics*, 32:154–156, 2015. doi: <https://doi.org/10.1093/bioinformatics/btv514>. URL <https://academic.oup.com/bioinformatics/article/32/1/154/1742487/ICT-isotope-correction-toolbox>.
- [88] J. Kalinowski, B. Bathe, D. Bartels, N. Bischoff, M. Bott, A. Burkovski, N. Dusch, L. Eggeling, B. J. Eikmanns, L. Gaigalat, A. Goesmann, M. Hartmann, K. Huthmacher, R. Krämer, B. Linke, A. C. McHardy, F. Meyer, B. Möckel, W. Pfefferle, A. Pühler, D. a. Rey, C. Rückert, O. Rupp, H. Sahm, V. F. Wendisch, I. Wiegräbe, and A. Tauch. The complete *Corynebacterium glutamicum* ATCC 13032 genome sequence and its impact on the production of L-aspartate-derived amino acids and vitamins. *Journal of Biotechnology*, 104(1-3):5–25, Sept. 2003. ISSN 01681656. URL <http://linkinghub.elsevier.com/retrieve/pii/S0168165603001548>.
- [89] B. Kanawati and P. Schmitt-Kopplin. Exploring rearrangements along the fragmentation of glutaric acid negative ion: a combined experimental and theoretical study. *Rapid Communications in Mass Spectrometry*, 24:1198–1206, 2010. doi: [doi:10.1002/rcm.4504](https://doi.org/10.1002/rcm.4504). URL <http://onlinelibrary.wiley.com/doi/10.1002/rcm.4504/full>.
- [90] J. Kappelmann, W. Wiechert, and S. Noack. Cutting the Gordian Knot: Identifiability of anaplerotic reactions in *Corynebacterium glutamicum* by means of ^{13}C -metabolic flux analysis. *Biotechnology and Bioengineering*, 113(3):661–674, 2016. ISSN 1097-0290. doi: [10.1002/bit.25833](https://doi.org/10.1002/bit.25833). URL <http://dx.doi.org/10.1002/bit.25833>.
- [91] J. Kappelmann, B. Klein, P. Geilenkirchen, and S. Noack. Comprehensive and accurate tracking of carbon origin of LC-tandem mass spectrometry collisional fragments for ^{13}C -MFA. *Analytical and Bioanalytical Chemistry*, 409(9):2309–2326, 2017. ISSN 1618-2650. doi: [10.1007/s00216-016-0174-9](https://doi.org/10.1007/s00216-016-0174-9). URL <http://dx.doi.org/10.1007/s00216-016-0174-9>.
- [92] Y. Karpievitch, J. Stanley, T. Taverner, J. Huang, J. N. Adkins, C. Ansong, and A. R. Dabney. A statistical framework for protein quantitation in bottom-up MS-based proteomics. *Bioinformatics*, 25:2028–2034, 2009. doi: [10.1093/bioinformatics/btp362](https://doi.org/10.1093/bioinformatics/btp362). URL <http://doi.org/10.1093/bioinformatics/btp362>.
- [93] P. Kebarle and L. Tang. From ions in solution to ions in the gas phase - the mechanism of electrospray mass spectrometry. *Analytical Chemistry*, 65(22):972A–986A, 1993. doi: [10.1021/ac00070a001](https://doi.org/10.1021/ac00070a001). URL <http://dx.doi.org/10.1021/ac00070a001>.
- [94] P. Kebarle and U. H. Verkerk. Electrospray: From ions in solution to ions in the gas phase, what we know now. *Mass Spectrometry Reviews*, 28(6):898–917, 2009. ISSN 1098-2787. doi: [10.1002/mas.20247](https://doi.org/10.1002/mas.20247). URL <http://dx.doi.org/10.1002/mas.20247>.

- [95] P. Kiefer, C. Nicolas, F. Letisse, and J.-C. Portais. Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. *Analytical Biochemistry*, 360(2):182 – 188, 2007. ISSN 0003-2697. doi: <http://dx.doi.org/10.1016/j.ab.2006.06.032>. URL <http://www.sciencedirect.com/science/article/pii/S0003269706004684>.
- [96] K. R. Kjeldsen and J. Nielsen. In silico genome-scale reconstruction and validation of the *Corynebacterium glutamicum* metabolic network. *Biotechnology and Bioengineering*, 102(2): 583–597, 2009. ISSN 1097-0290. doi: 10.1002/bit.22067. URL <http://dx.doi.org/10.1002/bit.22067>.
- [97] S. Klaffl and B. J. Eikmanns. Genetic and functional analysis of the soluble oxaloacetate decarboxylase from *Corynebacterium glutamicum*. *Journal of Bacteriology*, 192(10):2604–12, 2010. doi: 10.1128/JB.01678-09. URL <http://jb.asm.org/content/192/10/2604.abstract>.
- [98] M. I. Klapa, J. C. Aon, and G. Stephanopoulos. Systematic quantification of complex metabolic flux networks using stable isotopes and mass spectrometry. *European Journal of Biochemistry*, 270(17):3525–3542, 2003. ISSN 1432-1033. doi: 10.1046/j.1432-1033.2003.03732.x. URL <http://dx.doi.org/10.1046/j.1432-1033.2003.03732.x>.
- [99] S. Klein and E. Heinzle. Isotope labeling experiments in metabolomics and fluxomics. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, 4(3):261–272, 2012. ISSN 1939-005X. doi: 10.1002/wsbm.1167. URL <http://dx.doi.org/10.1002/wsbm.1167>.
- [100] T. Klein, K. Schneider, and E. Heinzle. A system of miniaturized stirred bioreactors for parallel continuous cultivation of yeast with online measurement of dissolved oxygen and off-gas. *Biotechnology and Bioengineering*, 110(2):535–542, 2013. ISSN 1097-0290. doi: 10.1002/bit.24633. URL <http://dx.doi.org/10.1002/bit.24633>.
- [101] A. Klingner, A. Bartsch, M. Dogs, I. Wagner-Döbler, D. Jahn, M. Simon, T. Brinkhoff, J. Becker, and C. Wittmann. Large-scale ^{13}C -flux profiling reveals conservation of the Entner-Doudoroff pathway as glycolytic strategy among glucose-using marine bacteria. *Applied and Environmental Microbiology*, 2015. doi: 10.1128/AEM.03157-14. URL <http://aem.asm.org/content/early/2015/01/20/AEM.03157-14.abstract>.
- [102] L. Konermann, E. Ahadi, A. D. Rodriguez, and S. Vahidi. Unraveling the mechanism of electrospray ionization. *Analytical Chemistry*, 85(1):2–9, 2013. doi: 10.1021/ac302789c. URL <http://dx.doi.org/10.1021/ac302789c>. PMID: 23134552.
- [103] H. L. Kornberg. Anaplerotic Sequences in Microbial Metabolism. *Angewandte Chemie International Edition in English*, 4(7):558–565, 1965. doi: 10.1002/anie.196505581. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.196505581>.

- [104] J. P. Krause, T. Polen, J. W. Youn, D. Emer, B. J. Eikmanns, and V. F. Wendisch. Regulation of the malic enzyme gene *malE* by the transcriptional regulator MalR in *Corynebacterium glutamicum*. *Journal of Biotechnology*, 159(3):204–15, 2012. ISSN 0168-1656. doi: <http://dx.doi.org/10.1016/j.jbiotec.2012.01.003>. URL <http://www.sciencedirect.com/science/article/pii/S0168165612000211>.
- [105] H. L. Kronberg and H. Krebs. Synthesis of Cell Constituents from C2-Units by a Modified Tricarboxylic Acid Cycle. *Nature*, 179:988–991, 1957. doi: 10.1038/179988a0. URL <https://www.nature.com/articles/179988a0>.
- [106] A. Laplante, B. Comte, and C. Desrosiers. Assay of blood and tissue oxaloacetate and α -ketoglutarate by isotope dilution gas chromatography-mass spectrometry. *Analytical Biochemistry*, 224(2):580 – 587, 1995. ISSN 0003-2697. doi: <http://dx.doi.org/10.1006/abio.1995.1090>. URL <http://www.sciencedirect.com/science/article/pii/S0003269785710901>.
- [107] H.-S. Lee and B.-J. Hwang. Methionine biosynthesis and its regulation in *Corynebacterium glutamicum*: parallel pathways of transsulfuration and direct sulphydrylation. *Applied Microbiology and Biotechnology*, 62(5):459–467, 2003. doi: 10.1007/s00253-003-1306-7. URL <https://doi.org/10.1007/s00253-003-1306-7>.
- [108] R. W. Leighty and M. R. Antoniewicz. Parallel labeling experiments with [U- ^{13}C]glucose validate *E. coli* metabolic network model for ^{13}C metabolic flux analysis. *Metabolic Engineering*, 14(5): 533 – 541, 2012. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2012.06.003>. URL <http://www.sciencedirect.com/science/article/pii/S1096717612000687>.
- [109] R. W. Leighty and M. R. Antoniewicz. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. *Metabolic Engineering*, 20(0):49 – 55, 2013. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2013.08.006>. URL <http://www.sciencedirect.com/science/article/pii/S1096717613000840>.
- [110] D. C. Liebler and L. J. Zimmerman. Targeted Quantitation of Proteins by Mass Spectrometry. *Biochemistry*, 52(22):3797–3806, 2013. doi: 10.1021/bi400110b. URL <http://dx.doi.org/10.1021/bi400110b>.
- [111] T. Loerting and J. Bernard. Aqueous Carbonic Acid (H_2CO_3). *ChemPhysChem*, 11(11):2305–2309, 2010. ISSN 1439-7641. doi: 10.1002/cphc.201000220. URL <http://dx.doi.org/10.1002/cphc.201000220>.
- [112] C. P. Long and M. R. Antoniewicz. Quantifying Biomass Composition by Gas Chromatography/Mass Spectrometry. *Analytical Chemistry*, 86(19):9423–9427, 2014. doi: 10.1021/ac502734e. URL <http://dx.doi.org/10.1021/ac502734e>.

- [113] F. Lottspeich and J. W. Engels. *Bioanalytik (third edition)*. Spektrum Akademischer Verlag, Heidelberg, 2012.
- [114] B. Luo, K. Groenke, R. Takors, C. Wandrey, and M. Oldiges. Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by liquid chromatography-mass spectrometry. *Journal of Chromatography A*, 1147: 153–164, 2007. ISSN 0021-9673. doi: <http://dx.doi.org/10.1016/j.chroma.2007.02.034>. URL <http://www.sciencedirect.com/science/article/pii/S0021967307002907>.
- [115] T. Mairinger, M. Steiger, J. Nocon, D. Mattanovich, G. Koellensperger, and S. Hann. Gas Chromatography-Quadrupole Time-of-Flight Mass Spectrometry-Based Determination of Isotopologue and Tandem Mass Isotopomer Fractions of Primary Metabolites for ^{13}C -Metabolic Flux Analysis. *Analytical Chemistry*, 87(23):11792–11802, 2015. doi: 10.1021/acs.analchem.5b03173. URL <http://dx.doi.org/10.1021/acs.analchem.5b03173>.
- [116] A. Martzolf, E. Cahoreau, G. Cogne, L. Peyriga, J.-C. Portais, E. Dechandol, F. Le Grand, S. Massou, O. Gonçalves, J. Pruvost, and J. Legrand. Photobioreactor design for isotopic non-stationary ^{13}C -metabolic flux analysis (INST ^{13}C -MFA) under photoautotrophic conditions. *Biotechnology and Bioengineering*, 109(12):3030–3040, 2012. ISSN 1097-0290. doi: 10.1002/bit.24575. URL <http://dx.doi.org/10.1002/bit.24575>.
- [117] A. Marx, A. A. De Graaf, W. Wiechert, and L. Eggeling. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. *Biotechnology and Bioengineering*, 49:11–129, 1996. ISSN 1097-0290. doi: 10.1002/(SICI)1097-0290(19960120)49:2<111::AID-BIT1>3.0.CO;2-T. URL [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(19960120\)49:2<111::AID-BIT1>3.0.CO;2-T](http://dx.doi.org/10.1002/(SICI)1097-0290(19960120)49:2<111::AID-BIT1>3.0.CO;2-T).
- [118] MATLAB. (*R2013b*). The MathWorks Inc., Natick, Massachusetts, 2013.
- [119] B. K. Matuszewski, M. L. Constanzer, and C. M. Chavez-Eng. Strategies for the Assessment of Matrix Effect in Quantitative Bioanalytical Methods Based on HPLC-MS/MS. *Analytical Chemistry*, 75(13):3019–3030, 2003. doi: 10.1021/ac020361s. URL <http://dx.doi.org/10.1021/ac020361s>. PMID: 12964746.
- [120] D. McCloskey, J. D. Young, S. Xu, B. O. Palsson, and A. M. Feist. MID Max: LC-MS/MS Method for Measuring the Precursor and Product Mass Isotopomer Distributions of Metabolic Intermediates and Cofactors for Metabolic Flux Analysis Applications. *Analytical Chemistry*, 88(2): 1362–1370, 2016. doi: 10.1021/acs.analchem.5b03887. URL <http://dx.doi.org/10.1021/acs.analchem.5b03887>. PMID: 26666286.
- [121] P. McCullagh and J. A. Nelder. *Generalized Linear Models, Second Edition*. Chapman and Hall/CRC, London, 1989.

- [122] F. W. McLafferty and F. Turecek. *Interpretation of Mass Spectra, Fourth Edition*. University Science Books, Sausalito, California, 1993.
- [123] J. Meija and J. A. Caruso. Deconvolution of isobaric interferences in mass spectra. *Journal of the American Society for Mass Spectrometry*, 15(5):654 – 658, 2004. ISSN 1044-0305. doi: <http://dx.doi.org/10.1016/j.jasms.2003.12.016>. URL <http://www.sciencedirect.com/science/article/pii/S1044030504000169>.
- [124] S. M. Miladinović, L. Fornelli, Y. Lu, K. M. Piech, H. H. Girault, and Y. O. Tsybin. In-spray supercharging of peptides and proteins in electrospray ionization mass spectrometry. *Analytical Chemistry*, 84(11):4647–4651, 2012. doi: 10.1021/ac300845n. URL <http://dx.doi.org/10.1021/ac300845n>. PMID: 22571167.
- [125] P. Millard, S. Massou, J.-C. Portais, and F. Létisse. Isotopic Studies of Metabolic Systems by Mass Spectrometry: Using Pascals Triangle To Produce Biological Standards with Fully Controlled Labeling Patterns. *Analytical Chemistry*, 86(20):10288–10295, 2014. doi: 10.1021/ac502490g. URL <http://dx.doi.org/10.1021/ac502490g>.
- [126] P. Millard, S. Massou, C. Wittmann, J.-C. Portais, and F. Létisse. Sampling of intracellular metabolites for stationary and non-stationary ^{13}C metabolic flux analysis in *Escherichia coli*. *Analytical Biochemistry*, 465:38 – 49, 2014. ISSN 0003-2697. doi: <http://dx.doi.org/10.1016/j.ab.2014.07.026>. URL <http://www.sciencedirect.com/science/article/pii/S0003269714003285>.
- [127] M. Möllney, W. Wiechert, D. Kownatzki, and A. A. de Graaf. Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments. *Biotechnology and Bioengineering*, 66(2):86–103, 1999. ISSN 1097-0290. doi: 10.1002/(SICI)1097-0290(1999)66:2<86::AID-BIT2>3.0.CO;2-A. URL [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(1999\)66:2<86::AID-BIT2>3.0.CO;2-A](http://dx.doi.org/10.1002/(SICI)1097-0290(1999)66:2<86::AID-BIT2>3.0.CO;2-A).
- [128] K. K. Murray, R. K. Boyd, M. N. Eberlin, G. J. Langley, L. Li, and Y. Naito. Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013). *Pure and Applied Chemistry*, 85(7):1515–1609, 2013. URL <http://dx.doi.org/10.1351/PAC-REC-06-04-06>.
- [129] F. Neidhardt, J. Ingraham, and M. Schaechter. *Physiology of the bacterial cell. A molecular approach*. Sinauer associates, Sunderland, MA, 1990.
- [130] K. Nöh, A. Wahl, and W. Wiechert. Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. *Metabolic Engineering*, 8(6):554 – 577, 2006. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2006.05.006>. URL <http://www.sciencedirect.com/science/article/pii/S1096717606000449>.

- [131] A. Nicolae, J. Wahrheit, J. Bahnemann, A.-P. Zeng, and E. Heinzle. Non-stationary ^{13}C metabolic flux analysis of Chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation. *BMC Systems Biology*, 8(1):50, 2014. ISSN 1752-0509. doi: 10.1186/1752-0509-8-50. URL <http://www.biomedcentral.com/1752-0509/8/50>.
- [132] S. Nidenführ, W. Wiechert, and K. Nöh. How to measure metabolic fluxes: a taxonomic guide for ^{13}C fluxomics. *Current Opinion in Biotechnology*, 34:82 – 90, 2015. ISSN 0958-1669. doi: <http://dx.doi.org/10.1016/j.copbio.2014.12.003>. URL <http://www.sciencedirect.com/science/article/pii/S0958166914002183>.
- [133] S. Noack and W. Wiechert. Quantitative metabolomics: a phantom? *Trends in Biotechnology*, 32(5):238 – 244, 2014. ISSN 0167-7799. doi: <http://dx.doi.org/10.1016/j.tibtech.2014.03.006>. URL <http://www.sciencedirect.com/science/article/pii/S0167779914000560>.
- [134] S. Noack, K. Nöh, M. Moch, M. Oldiges, and W. Wiechert. Stationary versus non-stationary ^{13}C -MFA: A comparison using a consistent dataset. *Journal of Biotechnology*, 154(2–3):179 – 190, 2011. ISSN 0168-1656. doi: <http://dx.doi.org/10.1016/j.jbiotec.2010.07.008>. URL <http://www.sciencedirect.com/science/article/pii/S0168165610003263>.
- [135] K. Nöh, K. Grönke, B. Luo, R. Takors, M. Oldiges, and W. Wiechert. Metabolic flux analysis at ultra short time scale: Isotopically non-stationary ^{13}C labeling experiments. *Journal of Biotechnology*, 129(2):249 – 267, 2007. ISSN 0168-1656. doi: <http://dx.doi.org/10.1016/j.jbiotec.2006.11.015>. URL <http://www.sciencedirect.com/science/article/pii/S0168165606009722>.
- [136] N. Okahashi, S. Kawana, J. Iida, H. Shimizu, and F. Matsuda. GC-MS/MS survey of collision-induced dissociation of *tert*-butyldimethylsilyl-derivatized amino acids and its application to ^{13}C -metabolic flux analysis of *Escherichia coli* central metabolism. *Analytical and Bioanalytical Chemistry*, 408(22):6133–6140, 2016. ISSN 1618-2650. doi: 10.1007/s00216-016-9724-4. URL <http://dx.doi.org/10.1007/s00216-016-9724-4>.
- [137] S. Okino, R. Noburyu, M. Suda, T. Jojima, M. Inui, and H. Yukawa. An efficient succinic acid production process in a metabolically engineered *Corynebacterium glutamicum* strain. *Applied Microbiology and Biotechnology*, 81(3):459–464, 2008. doi: 10.1007/s00253-008-1668-y. URL <https://doi.org/10.1007/s00253-008-1668-y>.
- [138] M. Oldiges, B. J. Eikmanns, and B. Blombach. Application of metabolic engineering for the biotechnological production of L-valine. *Applied Microbiology and Biotechnology*, 98(13):5859–5870, 2014. ISSN 0175-7598. doi: 10.1007/s00253-014-5782-8. URL <http://dx.doi.org/10.1007/s00253-014-5782-8>.

- [139] L. Ouerdane, J. Meija, S. Bakirdere, Y. L., and M. Z. Nonlinear signal response in electrospray mass spectrometry: implications for quantitation of arsenobetaine using stable isotope labeling by liquid chromatography and electrospray Orbitrap mass spectrometry. *Analytical Chemistry*, 84:3958–64, 2012. doi: 10.1021/ac203137n. URL <http://dx.doi.org/10.1021/ac203137n>.
- [140] S. M. Park, C. Shaw-Reid, A. J. Sinskey, and G. Stephanopoulos. Elucidation of anaplerotic pathways in *Corynebacterium glutamicum* via ^{13}C -NMR spectroscopy and GC-MS. *Applied Microbiology and Biotechnology*, 47(4):430–440, 1997. ISSN 1432-0614. doi: 10.1007/s002530050952. URL <http://dx.doi.org/10.1007/s002530050952>.
- [141] S.-Y. Park, M.-W. Moon, B. Subhadra, and J.-K. Lee. Functional characterization of the glxR deletion mutant of *Corynebacterium glutamicum* ATCC 13032: involvement of GlxR in acetate metabolism and carbon catabolite repression. *FEMS Microbiology Letters*, 304(2):107–115, 2010. doi: 10.1111/j.1574-6968.2009.01884.x. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1574-6968.2009.01884.x>.
- [142] J. Pauling, R. Röttger, A. Tauch, V. Azevedo, and J. Baumbach. CoryneRegNet 6.0—Updated database content, new analysis methods and novel features focusing on community demands. *Nucleic Acids Research*, 40(D1):D610, 2011. doi: 10.1093/nar/gkr883. URL <https://doi.org/10.1093/nar/gkr883>.
- [143] P. G. Peters-Wendisch, C. Kreutzer, J. Kalinowski, M. Patek, H. Sahm, and B. J. Eikmanns. Pyruvate carboxylase from *Corynebacterium glutamicum* : characterization , expression and inactivation of the pyc gene. *Microbiology*, 144:915–927, 1998. URL <http://mic.microbiologyresearch.org/content/journal/micro/10.1099/00221287-144-4-915>.
- [144] S. Petersen, A. A. de Graaf, L. Eggeling, M. Möllney, W. Wiechert, and H. Sahm. *In Vivo* Quantification of Parallel and Bidirectional Fluxes in the Anaplerosis of *Corynebacterium glutamicum*. *The Journal of biological chemistry*, 275(46):35932–41, 2000. doi: 10.1074/jbc.M908728199. URL <http://www.jbc.org/content/275/46/35932.abstract>.
- [145] S. Petersen, C. Mack, A. A. De Graaf, C. Riedel, B. J. Eikmanns, and H. Sahm. Metabolic Consequences of Altered Phosphoenolpyruvate Carboxykinase Activity in *Corynebacterium glutamicum* Reveal Anaplerotic Regulation Mechanisms *in Vivo*. *Metabolic Engineering*, 3(4): 344–61, 2001. ISSN 1096-7176. doi: <http://dx.doi.org/10.1006/mben.2001.0198>. URL <http://www.sciencedirect.com/science/article/pii/S1096717601901983>.
- [146] K. Pfeifer-Sancar, A. Mentz, C. Rückert, and J. Kalinowski. Comprehensive analysis of the *Corynebacterium glutamicum* transcriptome using an improved RNAseq technique. *BMC Genomics*, 14(1):1–23, 2013. ISSN 1471-2164. doi: 10.1186/1471-2164-14-888. URL <http://dx.doi.org/10.1186/1471-2164-14-888>.

- [147] P. Picotto and R. Aebersold. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods*, 9(6):555–566, 2012. doi: 10.1038/nmeth.2015. URL <http://www.nature.com/nmeth/journal/v9/n6/full/nmeth.2015.html>.
- [148] M. Piraud, C. Vianey-Saban, K. Petritis, C. Elfakir, J.-P. Steghens, A. Morla, and D. Bouchu. ESI-MS/MS analysis of underivatized amino acids: a new tool for the diagnosis of inherited disorders of amino acid metabolism. Fragmentation study of 79 molecules of biological interest in positive and negative ionisation mode. *Rapid Communications in Mass Spectrometry*, 17(12):1297–1311, 2003. ISSN 1097-0231. doi: 10.1002/rcm.1054. URL <http://dx.doi.org/10.1002/rcm.1054>.
- [149] J. Plassmeier, M. Persicke, A. Pühler, C. Sterthoff, C. Rückert, and J. Kalinowski. Molecular characterization of PrpR, the transcriptional activator of propionate catabolism in *Corynebacterium glutamicum*. *Journal of Biotechnology*, 159(1):1 – 11, 2012. ISSN 0168-1656. doi: <https://doi.org/10.1016/j.jbiotec.2011.09.009>. URL <http://www.sciencedirect.com/science/article/pii/S0168165611005451>.
- [150] V. Puech, N. Bayan, K. Salim, G. Leblon, and M. Daffé. Characterization of the *in vivo* acceptors of the mycoloyl residues transferred by the corynebacterial PS1 and the related mycobacterial antigens 85. *Molecular Microbiology*, 35(5):1026–1041, 2000. ISSN 1365-2958. doi: 10.1046/j.1365-2958.2000.01738.x. URL <http://dx.doi.org/10.1046/j.1365-2958.2000.01738.x>.
- [151] L.-E. Quek, C. Wittmann, L. K. Nielsen, and J. O. Krömer. OpenFLUX: efficient modelling software for ^{13}C -based metabolic flux analysis. *Microbial Cell Factories*, 8(1):25, 2009. ISSN 1475-2859. doi: 10.1186/1475-2859-8-25. URL <http://dx.doi.org/10.1186/1475-2859-8-25>.
- [152] L. Ramaley and L. C. Herrera. Software for the calculation of isotope patterns in tandem mass spectrometry. *Rapid Communications in Mass Spectrometry*, 22(17):2707–2714, 2008. ISSN 1097-0231. doi: 10.1002/rcm.3668. URL <http://dx.doi.org/10.1002/rcm.3668>.
- [153] A. Repta and T. Higuchi. Synthesis, isolation, and some chemistry of citric acid anhydride. *Journal of Pharmaceutical Sciences*, 58(9):1110 – 1114, 1969. ISSN 0022-3549. doi: <http://dx.doi.org/10.1002/jps.2600580915>. URL <http://www.sciencedirect.com/science/article/pii/S0022354915369926>.
- [154] C. Riedel, D. Rittmann, P. Dangel, B. Moeckel, S. Petersen, H. Sahm, and B. J. Eikmanns. Characterization of the phosphoenolpyruvate carboxykinase gene from *Corynebacterium glutamicum* and significance of the enzyme for growth and amino acid production. *Journal of Molecular Microbiology and Biotechnology*, 3(4):573–83, 2001.
- [155] F. Rogalewicz, Y. Hoppilliard, and G. Ohanessian. Fragmentation mechanisms of α -amino acids protonated under electrospray ionization: a collisional activation and ab initio theoretical study. *International Journal of Mass Spectrometry*, 196(0):565 – 590, 2000. ISSN 1387-3806.

- doi: [http://dx.doi.org/10.1016/S1387-3806\(99\)00225-0](http://dx.doi.org/10.1016/S1387-3806(99)00225-0). URL <http://www.sciencedirect.com/science/article/pii/S1387380699002250>.
- [156] M. Rühl, B. Rupp, K. Nöh, W. Wiechert, U. Sauer, and N. Zamboni. Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of ^{13}C metabolic flux analysis. *Biotechnology and Bioengineering*, 109(3):763–771, 2012. ISSN 1097-0290. doi: 10.1002/bit.24344. URL <http://dx.doi.org/10.1002/bit.24344>.
- [157] R. Sander. Compilation of Henry’s law constants (version 4.0) for water as solvent. *Atmospheric Chemistry and Physics*, 15(8):4399–4981, 2015. doi: 10.5194/acp-15-4399-2015. URL <http://www.atmos-chem-phys.net/15/4399/2015/>.
- [158] U. Sauer and B. Eikmanns. The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiology Reviews*, 29(4):765 – 794, 2005. ISSN 1574-6976. doi: 10.1016/j.femsre.2004.11.002. URL <http://dx.doi.org/10.1016/j.femsre.2004.11.002>.
- [159] A. Schmidt, K. Kochanowski, S. Vedelar, E. Ahrne, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann. The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology*, 34:104–110, 2016. doi: 10.1038/nbt.3418. URL <http://www.nature.com/nbt/journal/v34/n1/full/nbt.3418.html>.
- [160] K. Schmidt, M. Carlsen, J. Nielsen, and J. Villadsen. Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnology and Bioengineering*, 55(6):831–840, 1997. ISSN 1097-0290. doi: 10.1002/(SICI)1097-0290(19970920)55:6<831::AID-BIT2>3.0.CO;2-H. URL [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(19970920\)55:6<831::AID-BIT2>3.0.CO;2-H](http://dx.doi.org/10.1002/(SICI)1097-0290(19970920)55:6<831::AID-BIT2>3.0.CO;2-H).
- [161] O. T. Schubert, L. C. Gillet, B. C. Collins, P. Navarro, G. Rosenberger, W. E. Wolski, H. Lam, D. Amodei, P. Mallick, B. MacLean, and R. Aebersold. Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nature Protocols*, 10:426–441, 2015. doi: 10.1038/nprot.2015.015. URL <https://www.nature.com/articles/nprot.2015.015>.
- [162] Y. Shiio and R. Aebersold. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nature Protocols*, 1(1):139–145, 2006. doi: 10.1038/nprot.2006.22. URL <https://www.nature.com/articles/nprot.2006.22>.
- [163] T. Shirai, K. Fujimura, C. Furusawa, K. Nagahisa, S. Shioya, and H. Shimizu. Study on roles of anaplerotic pathways in glutamate overproduction of *Corynebacterium glutamicum* by metabolic flux analysis. *Microbial Cell Factories*, 6(1):19, Jun 2007. ISSN 1475-2859. doi: 10.1186/1475-2859-6-19. URL <https://doi.org/10.1186/1475-2859-6-19>.

- [164] D. Siegel, H. Permentier, D.-J. Reijngoud, and R. Bischoff. Chemical and technical challenges in the analysis of central carbon metabolites by liquid-chromatography mass spectrometry. *Journal of Chromatography B*, 966(0):21 – 33, 2014. ISSN 1570-0232. doi: <http://dx.doi.org/10.1016/j.jchromb.2013.11.022>. URL <http://www.sciencedirect.com/science/article/pii/S1570023213006351>.
- [165] *Sigma quality control test procedure Enzymatic Assay of Pyruvate Carboxylase*. Sigma Aldrich, 1999. URL https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/General_Information/pyruvate_carboxylase.pdf.
- [166] *Product Information Aconitase from porcine heart*. Sigma Aldrich, 2003. URL http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/a5384pis.pdf.
- [167] *Assay Procedure for L-Glutamic Dehydrogenase (NADP)*. Sigma Aldrich, 2015. URL <http://www.sigmaaldrich.com/technical-documents/protocols/biology/assay-procedure-for-l-glutamic-dehydrogenase-nadp.html> (2015-03-16).
- [168] L. Sleno and D. A. Volmer. Ion activation methods for tandem mass spectrometry. *Journal of Mass Spectrometry*, 39(10):1091–1112, 2004. ISSN 1096-9888. doi: 10.1002/jms.703. URL <http://dx.doi.org/10.1002/jms.703>.
- [169] R. M. Smith. *Understanding Mass Spectra: A Basic Approach, 2nd Edition*. John Wiley and Sons, Inc., 2004.
- [170] A. Teleki, A. Sánchez-Kopper, and R. Takors. Alkaline conditions in hydrophilic interaction liquid chromatography for intracellular metabolite quantification using tandem mass spectrometry. *Analytical Biochemistry*, 475:4 – 13, 2015. ISSN 0003-2697. doi: <http://dx.doi.org/10.1016/j.ab.2015.01.002>. URL <http://www.sciencedirect.com/science/article/pii/S0003269715000068>.
- [171] N. Tepper and T. Shlomi. An integrated computational approach for metabolic flux analysis coupled with inference of tandem-MS collisional fragments. *Bioinformatics*, 29(23):3045–3052, 2013. doi: 10.1093/bioinformatics/btt516. URL <http://bioinformatics.oxfordjournals.org/content/29/23/3045.abstract>.
- [172] H. Teramoto, M. Inui, and H. Yukawa. Transcriptional regulators of multiple genes involved in carbon metabolism in *Corynebacterium glutamicum*. *Journal of Biotechnology*, 154(2):114 – 125, 2011. ISSN 0168-1656. doi: <https://doi.org/10.1016/j.jbiotec.2011.01.016>. URL <http://www.sciencedirect.com/science/article/pii/S0168165611000605>.
- [173] M. Tesch, B. J. Eikmanns, A. A. de Graaf, and H. Sahm. Ammonia assimilation in *Corynebacterium glutamicum* and a glutamate dehydrogenase-deficient mutant. *Biotechnology Letters*,

- 20(10):953–957, Oct 1998. ISSN 1573-6776. doi: 10.1023/A:1005406126920. URL <https://doi.org/10.1023/A:1005406126920>.
- [174] B. Thiele, K. Füllner, N. Stein, M. Oldiges, A. Kuhn, and D. Hofmann. Analysis of amino acids without derivatization in barley extracts by LC-MS-MS. *Analytical and Bioanalytical Chemistry*, 391(7):2663–2672, 2008. ISSN 1618-2642. doi: 10.1007/s00216-008-2167-9. URL <http://dx.doi.org/10.1007/s00216-008-2167-9>.
- [175] E. M. Thurman and I. Ferrer. The isotopic mass defect: a tool for limiting molecular formulas by accurate mass. *Analytical and Bioanalytical Chemistry*, 397(7):2807–2816, 2010. ISSN 1618-2650. doi: 10.1007/s00216-010-3562-6. URL <http://dx.doi.org/10.1007/s00216-010-3562-6>.
- [176] J. Tillack, N. Paczia, K. Nöh, W. Wiechert, and S. Noack. Error Propagation Analysis for Quantitative Intracellular Metabolomics. *Metabolites*, 2(4):1012–1030, 2012. ISSN 2218-1989. doi: 10.3390/metabo2041012. URL <http://www.mdpi.com/2218-1989/2/4/1012>.
- [177] S. Unthan, A. Grünberger, J. van Ooyen, J. Gätgens, J. Heinrich, N. Paczia, W. Wiechert, D. Kohlheyer, and S. Noack. Beyond growth rate 0.6: What drives *Corynebacterium glutamicum* to higher growth rates in defined medium. *Biotechnology and Bioengineering*, 111(2):359–371, 2014. ISSN 1097-0290. doi: 10.1002/bit.25103. URL <http://dx.doi.org/10.1002/bit.25103>.
- [178] D. Uy, S. Delaunay, J.-M. Engasser, and J.-L. Goergen. A method for the determination of pyruvate carboxylase activity during the glutamic acid fermentation with *Corynebacterium glutamicum*. *Journal of Microbiological Methods*, 39(1):91 – 96, 1999. ISSN 0167-7012. doi: [https://doi.org/10.1016/S0167-7012\(99\)00104-9](https://doi.org/10.1016/S0167-7012(99)00104-9). URL <http://www.sciencedirect.com/science/article/pii/S0167701299001049>.
- [179] J. van Ooyen, S. Noack, M. Bott, A. Reth, and L. Eggeling. Improved L-lysine production with *Corynebacterium glutamicum* and systemic insight into citrate synthase flux and activity. *Biotechnology and Bioengineering*, 109(8):2070–81, 2012. ISSN 1097-0290. doi: 10.1002/bit.24486. URL <http://dx.doi.org/10.1002/bit.24486>.
- [180] W. A. van Winden, J. J. Heijnen, P. J. T. Verheijen, and J. Grievink. A priori analysis of metabolic flux identifiability from ^{13}C -labeling data. *Biotechnology and Bioengineering*, 74(6):505–516, 2001. doi: 10.1002/bit.1142. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.1142>.
- [181] W. A. Van Winden, C. Wittmann, E. Heinzle, and J. J. Heijnen. Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnology and Bioengineering*, 80(4):477–479, 2002. ISSN 1097-0290. doi: 10.1002/bit.10393. URL <http://dx.doi.org/10.1002/bit.10393>.

- [182] R. Voges. *Targeted quantification of central metabolic enzymes in Corynebacterium glutamicum*. PhD thesis, RWTH Aachen University, 2013.
- [183] R. Voges and S. Noack. Quantification of proteome dynamics in *Corynebacterium glutamicum* by ^{15}N -labeling and selected reaction monitoring. *Journal of Proteomics*, 75(9):2660–9, 2012. ISSN 1874-3919. doi: <http://dx.doi.org/10.1016/j.jprot.2012.03.020>. URL <http://www.sciencedirect.com/science/article/pii/S1874391912001534>.
- [184] M. Vogt, S. Haas, S. Klaffl, T. Polen, L. Eggeling, J. van Ooyen, and M. Bott. Pushing product formation to its limit: Metabolic engineering of *Corynebacterium glutamicum* for L-leucine overproduction. *Metabolic Engineering*, 22(0):40 – 52, 2014. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2013.12.001>. URL <http://www.sciencedirect.com/science/article/pii/S1096717613001201>.
- [185] M. Vogt, K. Krumbach, W.-G. Bang, J. van Ooyen, S. Noack, B. Klein, M. Bott, and L. Eggeling. The contest for precursors: channelling L-isoleucine synthesis in *Corynebacterium glutamicum* without byproduct formation. *Applied Microbiology and Biotechnology*, pages 1–10, 2014. ISSN 0175-7598. doi: 10.1007/s00253-014-6109-5. URL <http://dx.doi.org/10.1007/s00253-014-6109-5>.
- [186] M. Wang, G. Audi, A. Wapstra, F. Kondev, M. MacCormick, X. Xu, and B. Pfeiffer. The Ame2012 atomic mass evaluation. *Chinese Physics C*, 36(12):1603, 2012. URL <http://stacks.iop.org/1674-1137/36/i=12/a=003>.
- [187] M. Weitzel. *High performance algorithms for metabolic flux analysis*. PhD thesis, University of Siegen, Germany, 2010.
- [188] M. Weitzel, K. Nöh, T. Dalman, S. Niedenführ, B. Stute, and W. Wiechert. 13CFLUX2—high-performance software suite for ^{13}C -metabolic flux analysis. *Bioinformatics*, 29(1):143–5, 2013. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3530911&tool=pmcentrez&rendertype=abstract>.
- [189] V. Wendisch, A. de Graaf, H. Sahm, and B. Eikmanns. Quantitative Determination of Metabolic Fluxes during Coutilization of Two Carbon Sources: Comparative Analyses with *Corynebacterium glutamicum* during Growth on Acetate and/or Glucose. *Journal of Bacteriology*, 182(11):3033–3096, 2000. doi: 10.1128/JB.182.11.3088-3096.2000. URL <http://jb.asm.org/content/182/11/3088.abstract>.
- [190] V. F. Wendisch, M. Spies, D. J. Reinscheid, S. Schnicke, H. Sahm, and B. J. Eikmanns. Regulation of acetate metabolism in *Corynebacterium glutamicum*: transcriptional control of the isocitrate lyase and malate synthase genes. *Archives of Microbiology*, 168(4):262–269, 1997. ISSN 1432-072X. doi: 10.1007/s002030050497. URL <http://dx.doi.org/10.1007/s002030050497>.

- [191] F. C. Werner. *Wortelemente lateinisch-griechischer Fachausdrücke in den biologischen Wissenschaften*. suhrkamp taschenbuch 64, 1972.
- [192] W. Wiechert. *Bioinformatics: From Nucleic Acids and Proteins to Cell Metabolism*, chapter Algebraic methods for the analysis of redundancy and identifiability in metabolic ^{13}C labelling systems, page 169. Verlag Chemie, 1995.
- [193] W. Wiechert and A. A. de Graaf. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnology and Bioengineering*, 55(1):101–117, 1997. ISSN 1097-0290. doi: 10.1002/(SICI)1097-0290(19970705)55:1<101::AID-BIT12>3.0.CO;2-P. URL [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(19970705\)55:1<101::AID-BIT12>3.0.CO;2-P](http://dx.doi.org/10.1002/(SICI)1097-0290(19970705)55:1<101::AID-BIT12>3.0.CO;2-P).
- [194] W. Wiechert, C. Siefke, A. A. De Graaf, and A. Marx. Bidirectional reaction steps in metabolic networks: II. Flux Estimation and statistical analysis. *Biotechnology and Bioengineering*, 55(1): 118–35, 1997. ISSN 1097-0290. doi: 10.1002/(SICI)1097-0290(19970705)55:1<118::AID-BIT13>3.0.CO;2-I. URL [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(19970705\)55:1<118::AID-BIT13>3.0.CO;2-I](http://dx.doi.org/10.1002/(SICI)1097-0290(19970705)55:1<118::AID-BIT13>3.0.CO;2-I).
- [195] W. Wiechert, M. Möllney, N. Isermann, M. Wurzel, and A. A. de Graaf. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnology and Bioengineering*, 66(2):69–85, 1999. ISSN 1097-0290. doi: 10.1002/(SICI)1097-0290(1999)66:2<69::AID-BIT1>3.0.CO;2-6. URL [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(1999\)66:2<69::AID-BIT1>3.0.CO;2-6](http://dx.doi.org/10.1002/(SICI)1097-0290(1999)66:2<69::AID-BIT1>3.0.CO;2-6).
- [196] C. Wittmann. Fluxome analysis using GC-MS. *Microbial Cell Factories*, 6:6, 2207. doi: 10.1186/1475-2859-6-6. URL <https://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-6-6>.
- [197] C. Wittmann and E. Heinzle. Genealogy Profiling through Strain Improvement by Using Metabolic Network Analysis : Metabolic Flux Genealogy of Several Generations of Lysine-Producing Corynebacteria. *Applied and Environmental Microbiology*, 68(12):5843–5859, 2002. doi: 10.1128/AEM.68.12.5843. URL <http://aem.asm.org/content/68/12/5843.abstract>.
- [198] S. Wolf, S. Schmidt, M. Müller-Hannemann, and S. Neumann. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics*, 11(148), 2010. doi: 10.1186/1471-2105-11-148. URL <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-148>.
- [199] Wolfram Research. *Mathematica, Version 8.0*. Wolfram Research, Inc., Champaign, Illinois, 2010.

- [200] L. Wu, M. R. Mashego, J. C. van Dam, A. M. Proell, J. L. Vinke, C. Ras, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly ^{13}C -labeled cell extracts as internal standard. *Analytical Biochemistry*, 336(2):164 – 171, 2005. ISSN 0003-2697. doi: <https://doi.org/10.1016/j.ab.2004.09.001>. URL <http://www.sciencedirect.com/science/article/pii/S0003269704007237>.
- [201] T. H. Yang, C. Wittmann, and E. Heinzle. Respirometric ^{13}C flux analysis, Part I: Design, construction and validation of a novel multiple reactor system using on-line membrane inlet mass spectrometry. *Metabolic Engineering*, 8(5):417 – 431, 2006. ISSN 1096-7176. doi: <http://dx.doi.org/10.1016/j.ymben.2006.03.001>. URL <http://www.sciencedirect.com/science/article/pii/S109671760600022X>.
- [202] H. Yoo, M. R. Antoniewicz, G. Stephanopoulos, and J. K. Kelleher. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. *Journal of Biological Chemistry*, 283(30):20621–20627, 2008. doi: 10.1074/jbc.M706494200. URL <http://www.jbc.org/content/283/30/20621.abstract>.
- [203] J. D. Young, D. K. Allen, and J. A. Morgan. Isotopomer Measurement Techniques in Metabolic Flux Analysis II: Mass Spectrometry. In G. Sriram, editor, *Plant Metabolism*, volume 1083 of *Methods in Molecular Biology*, pages 85–108. Humana Press, 2014. ISBN 978-1-62703-660-3. doi: 10.1007/978-1-62703-661-0_7. URL http://dx.doi.org/10.1007/978-1-62703-661-0_7.
- [204] J. Yuan, B. D. Bennett, and J. D. Rabinowitz. Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nature Protocols*, 3(8):1328–1340, 2008. doi: 10.1038/nprot.2008.131. URL <http://www.nature.com/nprot/journal/v3/n8/full/nprot.2008.131.html>.
- [205] M. Yuan, S. B. Breitkopf, X. Yang, and J. M. Asara. A positive/negative ion-switching, targeted mass spectrometry-based metabolomics platform for bodily fluids, cells, and fresh and fixed tissue. *Nature Protocols*, 7:872–881, 2012. doi: 10.1038/nprot.2012.024. URL <http://www.nature.com/nprot/journal/v7/n5/full/nprot.2012.024.html>.
- [206] N. Zamboni, S. Fendt, M. Rühl, and U. Sauer. ^{13}C -based metabolic flux analysis. *Nature protocols*, 4(6):878–92, 2009. doi: 10.1038/nprot.2009.58. URL <http://www.nature.com/nprot/journal/v4/n6/full/nprot.2009.58.html>.
- [207] R. M. Zelle, J. C. Harrison, J. T. Pronk, and A. J. A. van Maris. Anaplerotic Role for Cytosolic Malic Enzyme in Engineered *Saccharomyces cerevisiae* Strains. *Applied and Environmental Microbiology*, 77(3):732 – 738, 2011. doi: 10.1128/AEM.02132-10. URL <http://aem.asm.org/content/77/3/732.abstract>.

BIBLIOGRAPHY

- [208] C. Zhang, J. Zhang, Z. Kang, G. Du, X. Yu, T. Wang, and J. Chen. Enhanced production of L-phenylalanine in *Corynebacterium glutamicum* due to the introduction of *Escherichia coli* wild-type gene *aroH*. *Journal of Industrial Microbiology and Biotechnology*, 40(6):643–651, 2013. ISSN 1367-5435. doi: 10.1007/s10295-013-1262-x. URL <http://dx.doi.org/10.1007/s10295-013-1262-x>.
- [209] M. Zimmermann, U. Sauer, and N. Zamboni. Quantification and mass isotopomer profiling of α -keto acids in central carbon metabolism. *Anal. Chem.*, 86(6):3232 – 3237, 2014. doi: 10.1021/ac500472c. URL <http://dx.doi.org/10.1021/ac500472c>.