THESIS ON NATURAL AND EXACT SCIENCES B143

Quantitative Proteomics of *Escherichia coli*: From Relative to Absolute Scale

LIISA ARIKE

TUT PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Chemical and Material Technology
Department of Food Processing

This dissertation was accepted for the defense of the degree of Doctor of Philosophy (Chemical and Material Sciences) on June 12th, 2012.

Supervisors: Senior Scientist Kaarel Adamberg, Department of Food Processing, Tallinn University of Technology (TUT)

Professor Raivo Vilu, Department of Chemistry, Tallinn University of Technology (TUT)

Senior Scientist Lauri Peil, Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, UK

Opponents: Dr. Matthias Selbach, Department of Cell Signalling and Mass Spectrometry, Max Delbrück Center for Molecular Medicine, Berlin, Germany

Dr. Jaanus Remme, Institute of Molecular and Cell Biology, University of Tartu

Defense of the thesis: November 22nd, 2012

DECLARATION: I hereby declare that this doctoral thesis, submitted for the doctoral degree at TUT, is my original investigation and achievement and has not been submitted for the defense of any academic degree elsewhere.

Liisa Arike

COLOPHON: This thesis was typeset with LaTeX 2ε using André Miede's *classicthesis* style with modifications by Ardo Illaste and David W. Schryer to conform with TUT style guidelines. The main font is Libertine. ZapfChancery is used for Latin text. Biolinum is used for *sans* serif text.

ISSN 1406-4723
ISBN 978-9949-23-379-3 (publication)
ISBN 978-9949-23-380-9 (PDF)

European Union
European Social Fund
Investing in your future

LOODUS - JA TÄPPISTEADUSED B143

Kvantitatiivne *Escherichia coli* proteoomika: relatiivsetest numbritest absoluutsete kogusteni

LIISA ARIKE

TTÜ KIRJASTUS

CONTENTS

ABSTRACT

According to the most recent definition proteomics is a "large-scale study of endogenous proteins, their post-translational modifications, interactions and dynamic behavior in space and time" (Lamond et al., 2012). Quantitative proteomics has become a standard technique in molecular biology, next to transcriptomics and metabolomics, to measure cellular response to changing environmental conditions. There are several quantitative proteomics methods available and the development of new methods is an active research area. Proteome quantification can either be carried out in a relative way or on an absolute scale. Relative protein quantification methods allow comparison of relative protein abundances in samples and characterization of the proteome dynamics in cellular systems. However, absolute cellular protein concentrations are essential for quantitative and comprehensive understanding of organism's metabolism and for mathematical modelling in systems biology.

The objectives of this dissertation were as follows: 1) development of methods for the quantitative analysis of a growth rate dependent *Escherichia coli* proteome, 2) comparison of different quantification methods to characterize the *E. coli* proteome, and 3) to further our understanding of *E. coli* metabolism using proteome, transcriptome, metabolome and cultivation data.

Using an orthogonal acceleration time-of-flight mass spectrometer (oa-TOF-MS) we tested an "in-source" fragmentation method for identifying proteins. *E. coli* K-12 MG1655 cell lysate was separated by SDS-PAGE; fractions were digested and separated by ultra high performance liquid chromatography (UHPLC). Peptides were identified using oa-TOF-MS to measure exact masses of parent ions and the fragment ions generated by "in source" collision induced dissociation (CID). Fragmentation of all compounds was achieved by rapid cycling between high and low values of energy applied to the ions. More than 100 proteins of the *E. coli* K12 proteome were identified and relatively quantified. Results were found to correlate with transcriptome measured by DNA microarray (R^2 = 0.64). However, the up-front CID method was not able to comprehensively quantify enough of the proteome to study the metabolism of *E. coli*. Therefore, further studies were carried out with tandem mass-spectrometer LTQ Orbitrap.

Metabolic labeling with ^{15}N labeled ammonium salt ($^{15}NH_4Cl$) as the sole nitrogen source was used in a quantitative proteomics study of the growth rate dependent acetate overflow metabolism of *E. coli* K-12 MG1655. Metabolic labeling of growing cells is a reliable method for quantitative proteome measurements because the label is incorporated into the cells at the earliest stage possible. However, to reduce costs, ^{15}N labeled batch culture was used as a spike-in reference in this study. As a result, approximately 1,600 *E. coli* proteins were quantified at five different growth rates along with transcriptome data, resulting in a reasonable correlation (R^2 = 0.62-0.84 for biological replicates and R^2 = 0.51-0.62 for protein and mRNA comparison).

Although we obtained good coverage and biologically interesting results with relative quantification, many mathematical models applied in systems biology require absolute protein concentrations. Measuring concentrations in metabolome or clinical proteome studies is usually achieved using isotopically labeled standards. However, absolute quantification of the whole proteome by spiked-in isotope-labeled proteins, peptides or concatenated peptides is unfeasible due to its high cost and labor (Brownridge et al., 2011; Ludwig et al., 2012; Whiteaker et al., 2011). Label-free methods are an attractive alternative to labeling experiments. Absolute protein abundances were thus calculated using the APEX spectral counting method applied to the data from the ^{15}N labeled experiments. We demonstrated that from existing data it is possible to obtain absolute protein copy numbers per cell, but care must be taken in sample preparation because pre-fractionation of the samples might affect the accuracy of label-free methods.

To compare spectral counting with peak area measurement we carried out a shotgun experiment and calculated protein copies in the cell by spectral counting methods APEX and emPAI and by the intensity based method iBAQ. Good correlation (R^2 = 0.76-0.81) between the three methods was demonstrated. However, in the case of spectral counting methods, very large variance was observed for ribosomal proteins, and absolute protein concentrations were not normally distributed. Peak intensity based method iBAQ produced the best correlation between biological replicates, had a normal distribution of protein abundances in the cell, and had smallest variations for ribosomal proteins.

This dissertation forms the basis for further detailed studies of bacterial metabolism. Relative and absolute proteome quantification methods are now available as tools to be used in combination with transcriptomics, metabolomics, steady state cultivation, and metabolic modeling to elucidate quantitative peculiarities of bacterial physiology.

KOKKUVÕTE

Vastavalt kõige uuemale definitsioonile on proteoomika teadusharu, mis tegeleb kõigi organismis, organellis või rakus toodetud valkude ja nende post-translatoorsete modifikatsioonide uurimisega ajas ja ruumis (Lamond et al., 2012). Proteoomikast on saanud transkriptoomika ja metaboloomika kõrval oluline meetod molekulaarbioloogias mõistmaks raku või organismi vastust muutuvatele keskkonnatingimustele. Proteoomi kvantifitseerimiseks eksiteerib mitmeid meetodeid ning intensiivne uute meetodite välja töötamine on väga aktuaalne. Proteoomi kvantifitseerimist teostatakse suhtelisel või absoluutsel tasemel. Suhteline kvantifitseerimine võimaldab võrrelda sama valgu koguse erinevust proovide vahel ning valkude muutusi ajas. Samas ei anna suhteline kvantifitseerimine aimu kui palju molekule viib rakus läbi teatud protsesse. Absoluutne kvantifitseerimine võimaldab mõõta valkude kogust ühikutes ning võrrelda erinevate valkude konsentratsioone. Absoluutsed valkude kontsentratsioonid rakus on olulised süsteemibioloogias kasutatavates matemaatilistes modelleerimistes, et mõista terviklikult organismi ainevahetust.

Käesoleva uuringu eesmärgid olid järgmised: 1) kvantitatiivsete proteoomi mõõtmismeetodite juurutamine *Escherichia coli* kasvukiirusest sõltuva proteoomi uurimisel, 2) erinevate kvantitatiivsete proteoomi mõõtmismeetodite võrdlemine ja 3) pidevkultiveerimiskatsetest pärit kvantitatiivse proteoomikaandmete analüüs koos teiste –oomikaandmetega, et selgitada *E. coli* ainevahetuse iseärasusi.

Esmalt katsetasime uudset fragmenteerismismeetodit valkude identifitseerimiseks kasutades suhteliselt odavat lennuaja massispektromeetrit (TOF-MS). *E. coli* K-12 MG1655 rakulüsaat fraktsioneeriti geelelektroforeesi (SDS-PAGE) abil, valgu fraktsioonid lõigati ensüümiga trüpsiin peptiidideks ja saadud peptiidid lahutati ultra-kõrglahutus vedelikkromatograafia (UHPLC) abil. Peptiidid identifitseeriti nende täpse massi järgi ning seejärel lõhuti kõik mass-spektromeetrisse sisenenud ioonid fragmentideks, mis iseloomustasid peptiidide aminohappelist järjestust. Selline protsess erineb tavaliselt proteoomikas kasutatavast, sest: 1) kasutatakse ühekordset mass-spektromeetrit tavaliselt kasutatava tandem-mass-spektromeetri asemel; 2) lõhutakse kõik mass-spektromeetrisse sisenevad ioonid, samas kui tavaliselt lõhutakse 5-20 kõige intensiivsemat mitmekordselt laetud iooni. Antud meetodiga identifitseeriti ja kvantifitseeriti sadakond valku *E. coli* K-12 proteoomis, mis olid korrelatsioonis transkriptoomi analüüsi tulemustega (R^2 = 0,64). Kuna ei saavutatud piisavat proteoomi kattuvust, et uurida põhjalikult *E. coli* ainevahetuse iseärasusi, jätkati proteoomi analüüsi tandem-mass-spektromeetriga LTQ Orbitrap.

Kasvukiirusest sõltuva atsetaadi ülevoolu metabolismi uurimiseks *E. coli* K-12 MG1655 tüves proteoomi tasemel kastutati valkude metaboolset märgistamist, lisades söötmesse ^{15}N sisaldava ammooniumsoola ($^{15}NH_4Cl$). Metaboolne märgistamine on üks efektiivsemaid proteoomi kvantifitseerimise meetodeid, kuna isotoopne märgis lisatakse uuritavatesse valkudesse võimalikult varajases staadiumis ning edasine proovi töötlus ei põhjusta kvantifitseerimisse vigu. Märgistatud söötmete kasutamine pidevkultiveerimises on aga

väga kallis, seetõttu lisati pidevkultiveerimiskatsete proovidele standardina ^{15}N märgisega bakterikultuuri, mis oli kasvatatud eraldi perioodilise kutiveerimise eksperimendis. Kasutades SDS-PAGE fraktsioneerimist ja nano kõrglahutus vedelikkromatograafi koos LTQ Orbitrap mass-spektromeetriga identifitseeriti ja kvantifitseeriti suhtelisel skaalal umbes 1600 *E. coli* valku viiel erineval kasvukiirusel. Bioloogiliste replikaatide korrelatsioon (R^2 = 0,62-0,84) ja korrelatsioon transkriptoomi andmetega samadest katse punktidest (R^2 = 0,51-0,62) demonstreerisid head katsete reprodutseeritavust ning kontrollitud tingimustest tulenevat head valgu ja mRNA võrdlust.

Vaatamata sellele, et suhteline kvantifitseerimine andis hea proteoomi kattuvuse ning bioloogiliselt huvitavaid tulemusi, on süsteemibioloogia jaoks siiski vajalik ka valkude konsentratsioonide määramine. Biomarkerite tuvastamisel kasutatakse tihti märgistatud peptiide või valke, mida lisatakse proovile sisestandardina. Samas kogu *E. coli* proteoomi kvantifitseerimine tähendaks töötamist tuhande ja rohkema standardainega, see on tänu suurele töömahule ja maksumusele praegusel hetkel võimatu. Märgisevabad kvantifitseerimise meetodid on muutunud järjest populaarsemaks kui odav alternatiiv märgisega kvantifitseerimisele. Seetõttu arvutati eelnevalt metaboolselt märgistatud ja SDS-PAGE abil fraktsioneeritud proovide relatiivsetest andmetest MS spektrite loendamise tehnikaga (APEX) absoluutsed valkude kogused rakus. Kuna bioloogiliste replikaatide korratavus oli oodatust halvem, järeldasime, et olemasolevate andmete põhjal on küll võimalik arvutada absoluutsed valgumolekulide kogused proovis, kuid tuleb olla ettevaatlik proovi ettevalmistamisega, kuna eelnev fraktsioneerimine võib mõjutada märgisevaba kvantifitseerimise täpsust. Et kindel olla fraktsioneerimise negatiivses mõjus märgisevabale kvantifitseerimisele, teostasime samadele proovidele uue mass-spektromeetrilise analüüsi ilma SDS-PAGE fraktsioneerimiseta. Lisaks eelnevalt kasutatud spektrite lugemismeetodile APEX võtsime kasutusele ka teise spektrite loendamismeetodi emPAI ning piigi pindalal põhineva arvutusmeetodi iBAQ. Kasutatud kolm erinevat märgisevabalt valke kvantifitseerivat meetodit olid hästi korratavad bioloogilistele replikaatide puhul (R^2 = 0,89-0,99) ja tulemused korrelleerusid hästi ka omavahel: R^2 = 0,76-0,81. Samas avastati väga suur varieeruvus individuaalsete ribosomaalsete valkude kvantifitseerimisel (mis peaksid olema võrdsetes konsentratsioonides) just spektrite lugemismeetoditega APEX ja emPAI. Piigi pindalal põhinev meetod iBAQ andis parima reprodutseeritavuse bioloogilistele replikaatidele ning ribosomaalsete valkude konsentratsioonid erinesid üksteisest samuti kõige vähem.

Antud doktoritöös saadud tulemusi kasutatakse edasistel *E. coli* metabolismi kvantitatiivsetel uuringutel. Proteoomi kvantifitseerimismeetodid on nüüd kasutusele võetud ning rakendatakse koos transkriptoomika, metaboloomika ning pidevkultiveerimise ja modelleerimisega kvantitatiivsete rakufüsioloogia uuringutel ja uudsete tootjarakkude disainiprojektide läbiviimisel.

LIST OF PUBLICATIONS

The following publications form the basis of this dissertation and are reproduced in the appendices with permission from the publishers.

I Arike L, Valgepea K, Peil L, Nahku R, Adamberg K, Vilu R **Identification and relative quantification of proteins in *Escherichia coli* proteome by "up-front" collision-induced dissociation.** *European Journal of Mass Spectrometry*, 16(2):227-35 (2010)

II Valgepea K, Adamberg K, Nahku R, Lahtvee PJ, Arike L, Vilu R **Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase.** *BMC Systems Biology*, 4:166 (2010)

III Arike L, Valgepea, K, Peil, L, Nahku, R, Adamberg, K, Vilu, R **Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*.** *Journal of Proteomics*, 75(17):5437-5338 (2012)

SUMMARY OF AUTHOR'S CONTRIBUTION

The author assumed the main role setting up the proteomics methods applied at the Competence Centre of Food and Fermentation Technologies. This includes designing the relative and absolute quantification workflows for microbes and food samples.

I In Publication I, the author performed the experimental work, analysed and interpreted the data, and wrote the manuscript.

II In Publication II, the author designed and performed metabolic labelling, prepared the proteome samples, and analysed the proteome data.

III In Publication III, the author performed the experimental work, analysed and interpreted the data, and wrote the manuscript.

LIST OF PRESENTATIONS

I Arike L **Comparison of two common proteomics platform applied on growth rate dependent characterization of *Lactococcus lactis* proteome.** *Oral presentation at Waters 3rd Nordic User Meeting*, 11 September, 2012, Jurmala, Latvia.

II Arike L **Comparison and Applications of Label-free "Absolute" Proteome Quantification Methods on Study of Bacterial Proteome.** *Poster presentation at 60th Conference on Mass Spectrometry and Allied Topics, ASMS*, 20-24 May 2012, Vancouver, Canada.

III Arike L **From relative to absolute proteome quantification.** *Oral presentation at MaxQuant Summerschool*, 22-27 May 2011, Munich, Germany.

IV Arike L **Quantitative Proteomics Applied on Studies of Microorganisms.** *Oral presentation at 6th Joint Tartu – Turku – Tallinn Meeting "Exploring Science and Culture"*, 11-13 May 2011, Tallinn, Estonia.

V Arike L, Lahtvee, PJ, Valgepea, K, Nahku, R, Adamberg, K, Vilu R. **Characterization of Proteome Dynamics at Different Growth Rates in Continuous Cultures** *Poster presentation at Systems Biology of Microorganisms*, 22-24 March 2010, Paris, France.

VI Arike L, Valgepea K, Nahku R, Lahtvee PJ, Peil L, Adamberg K, Vilu R. **Quantitative study of *Escherichia coli* proteome by ^{15}N-labeling at different growth rates.** *Poster presentation at SPS Scientific Meeting: "Proteome Dynamics: Protein Quantification in Time and Space"*, 01-04 December 2009, Zurich, Switzerland.

VII Arike L, Nahku R, Lahtvee PJ, Adamberg K, Vilu R. **Identification and relative quantification of proteins in *Escherichia coli* proteome using up-front CID.** *Poster presentation at Proteomic Forum 2009*, 28 March - 02 April 2009, Berlin, Germany.

ADDITIONAL PUBLICATIONS

A Olspert A, Arike L, Peil L, Truve E. **Sobemovirus RNA linked to VPg over a threonine residue**. *FEBS Lett*, 585(19):2979-85 (2011).

B Lahtvee PJ, Adamberg K, Arike L, Nahku R, Aller, Vilu R. **Multi-omics approach to study the growth efficiency and amino acid metabolism in *Lactococcus lactis* at various specific growth rates**. *Microb Cell Fact*, 10:12 (2011).

C Nisamedtinov I, Kevvai K, Orumets K, Arike L, Sarand I, Korhola M, Paalme T. **Metabolic changes underlying the higher accumulation of glutathione in *Saccharomyces cerevisiae* mutants**. *Appl Microbiol Biotechnol*, 89(4):1029-37 (2011).

D Sumeri I, Arike L, Stekolštšikova J, Uusna R, Adamberg S, Adamberg K, Paalme T. **Effect of stress pretreatment on survival of probiotic bacteria in gastrointestinal tract simulator**. *Appl Microbiol Biotechnol*, 86(6):1925-31 (2010).

E Sumeri I, Arike L, Adamberg K, Paalme T. **Single bioreactor gastrointestinal tract simulator for study of survival of probiotic bacteria**. *Appl Microbiol Biotechnol*, 80(2):317-24 (2010).

ACKNOWLEDGMENTS

I am grateful to Prof. Raivo Vilu, who has always had great faith in me. He was the one who brought me to science when I was just a third year student. He was the one who directed me to mass spectrometry and proteomics when I did not have any clue of those techniques. And he is the one who always supports me if I have any doubts.

I am also thankful to Kaarel Adamberg for supervising me through last seven years. His deep knowledge in bacterial metabolism and physiology have been of great help. Probably I would not be still a scientist, if Kaarel would not been teaching me in the beginning. He made work in laboratory interesting and challenging.

My journey in proteomics would not have been so smooth if I would not found in the middle of the way my proteomics "guru" Lauri Peil. He is the one who I owe most of my skills in proteomics. Although he taught me in the hard way I value this experience very much. Without Lauri I would have kept on inventing bicycles.

Fellows — Kaspar, Karl, Petri, Ranno, Sten — it has been honour to be accepted in your team. I loved the wicked humour. Thank you for your support and constructive critics. I hope that my presence have been as useful for you as yours have been for me.

Girls — Kadri, Jana, Gethe — hang on.

Triin, Reet, Anett — thank you for teaching me teaching.

Andrus, Klim — without you I would be lost in data handling.

Mart — believe it or not, but my computer just does not work when you are not around.

All the colleagues in CCFFT, your support and help over the years means a lot to me and hopefully we will work together again one day.

My friends and family, who never have really understood what I am doing those long days in the laboratory, you have always cheered me up and never doubted in me.

Petri-Jaan, who is willing to discuss science with me 24/7 while also being wonderful partner in life. You have kept me going when I have been facing difficulties.

The financial support for this research was provided by the European Regional Development Fund project EU29994, SA Archimedes through the project 3.2.0701.11-0018 and Ministry of Education, Estonia, through the grant SF0140090s08. These studies were supported by European Social Fund's Doctoral Studies and Internationalization Programme DoRa. Programme DoRa is carried out by Archimedes Foundation. This work has been partially supported by graduate school "Functional materials and technologies" receiving funding from the European Social Fund under project 1.2.0401.09-0079 in Estonia.

ACRONYMS

2D	two dimensional
AIF	all-ion fragmentation
APEX	absolute protein expression
ATP	adenosine-5'-triphosphate
BSA	bovine serum albumin
CID	collision induced dissociation
COG	Cluster of Orthologous Groups
CV	coefficient of variation
DDA	data dependent analysis
DIA	data independent analysis
DNA	deoxyribonucleic acid
ECD	electron capture dissociation
emPAI	exponentially modified protein abundance index
ESI	electrospray ionization
ETD	electron transfer dissociation
FDR	false discovery rate
FT-ICR	Fourier transform ion cyclotron resonance
FWHM	full width at half maximum
HCD	higher energy collisional dissociation
HPLC	high performance liquid chromatography
iBAQ	intensity based absolute quantification
ICAT	isotope-coded affinity tag
IEF	isolectric focusing
IPG	immobilized pH gradient
iTRAQ	isobaric tags for relative and absolute quantitation
LC	liquid chromatography
MALDI	matrix-assisted laser desorption/ionization
MFA	metabolic flux analysis
mRNA	messenger RNA
MS	mass spectrometry
MS/MS	tandem mass spectrometry
m/z	mass to charge ratio
oa-TOF-MS	orthogonal acceleration time-of-flight mass spectrometry
ORF	open reading frame
PAI	protein abundance index
PQD	pulsed Q collision induced fragmentation
PTM	post-translational modification
Q-TOF	quadrupole time-of-flight
QqQ	triple quadrupole

RNA	ribonucleic acid
R_P	Pearson's correlation coefficient
R^2	squared Pearson's correlation coefficient
RP	reverse phase
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	stable isotope labeling by amino acids in cell culture
SRM	selected reaction monitoring
TIC	total ion count
TMT	tandem mass tag
TOF	time-of-flight
TOF-MS	time-of-flight mass spectrometry
UHPLC	ultra high performance liquid chromatography
UPS2	Sigma-Aldrich® universal proteomics standard

THESIS

1

INTRODUCTION

Systems biology takes advantage of transcriptome, metabolome, fluxome, and proteome measurements, in order to understand the regulation of the cellular metabolism (Nagaraj et al., 2011; Costenoble et al., 2011; Buescher et al., 2012). Recently, quantitative proteomics has become a standard procedure in molecular biology to measure cellular responses to changing environmental conditions (Walther and Mann, 2010). Proteome quantification can either be carried out in a relative way or on an absolute scale. While relative protein quantification methods allow one to compare protein ratios in samples and characterize proteome dynamics in cellular systems, absolute cellular protein concentrations are required for a quantitative understanding of metabolic processes and for mathematical modeling in systems biology. Knowledge of cellular protein concentrations enables one to evaluate the cost of running an active metabolic pathway or expressing enzymes in a stress response, estimate ribosomal translational capacity, calculate working rates of enzymes or metabolic capacity of cells, *etc.* Knowledge gained from these examples allows one to manipulate the metabolic behavior of the organism under study. For example, one can postpone or even preclude *Escherichia coli* acetate accumulation, a processes that is detrimental for target product synthesis, inhibits growth, and diverts valuable carbon from biomass formation (Nakano et al., 1997; Contiero et al., 2000).

The objective of this dissertation was to develop methods for the quantitative measurement of proteome and analysis of proteomics data together with the data of other omics' methods, in order to reveal quantitative features of the intracellular metabolism of microorganisms. Initially, a new method to analyze peptides on a single mass spectrometer was implemented. Because it resulted in a low number of protein identifications, more sophisticated equipment was used in further relative and absolute quantification analyses. Relative quantitative proteomics data, obtained by metabolic labeling with ^{15}N-enriched salt as the sole nitrogen source, was used to characterize mechanisms of acetate overflow in *E. coli* together with transcriptomics and metabolomics data. The absolute abundances of proteins were calculated from relative ratios by using a peak counting label-free method. Results were evaluated using a shotgun experiment to which internal standards were added. Based on the shotgun experiment, three label-free absolute quantification methods were compared and absolute proteome data was analyzed together with other omics data to characterize various regulatory mechanisms in the metabolism of *E. coli*.

2

LITERATURE REVIEW

2.1 PROTEOMICS

Proteomics is the study which aims to identify and quantify all proteins and their post-translational modifications, interactions and dynamic behavior in space and time, expressed in a cell, tissue or organism (Lamond et al., 2012). Although the idea of analysing the complete complement of proteins expressed in a sample arose already in 1970s, proteomics as a research area was defined in 1997 in order to make an analog with genomics (James, 1997).

2.1.1 *History of proteomics*

Although defined more than twenty years later the beginning of proteomics can be considered to be in 1970s, when Patrick H. O'Farrell managed to resolve 1,100 *Escherichia coli* proteins by two dimensional (2D) polyacrylamide gel by employing two independent properties of proteins to separate them: focusing by their isoelectric point in first dimension and separating them by sodium dodecyl sulfate (SDS) electrophoresis according their molecular weight in second dimension (O'Farrell, 1975). In order to identify proteins highly reproducible gels were required to map the gels and for identification purified proteins, mutants of known genes and specific antibodies were used (Neidhardt, 2011).

Protein identification remained slow and laborious until new methods emerged. In the 1970s Edman degradation (Edman, 1949) was automated, allowing one to sequence polypeptide chains faster by removing labelled N-terminal amino acids and identifying them by chromatography (Edman and Begg, 1967). A great improvement was the development of microsequencing techniques for electroblotted proteins (Aebersold et al., 1986, 1987) which allowed sequencing of less than 100 picomoles of protein.

Based on protein identifications 2D gel databases were established (for example human secreted proteins (Celis et al., 1987); *E. coli* K-12: (VanBogelen et al., 1990)) and identification of proteins became more straightforward. However, most of the reference maps contained only a small portion of proteins identified (Wilkins et al., 1996b). A key breakthrough with proteomics techniques arrived with the development of large biomolecule mass spectrometry analysis by soft ionization techniques MALDI (Tanaka et al., 1988; Karas and Hillenkamp, 1988) and ESI (Fenn et al., 1989). The inventors of both methods were rewarded with the Nobel prize in chemistry in 2002. Soft ionization techniques allowed one

to use mass spectrometry for protein identification and the first method to identify proteins from 2D gel spots was peptide mass fingerprinting (James et al., 1993; Pappin et al., 1993; Henzel et al., 1993). This technique is based on cleaving proteins into peptides using an enzyme and then determining the exact masses of the product peptides. These masses are then compared against *in silico* databases where protein sequences were cleaved with the same enzyme used in the experiment (Henzel et al., 1993). Peptide mass fingerprinting is suitable for analyzing single proteins or simple mixtures where one protein dominates. In order to analyze complex protein mixtures, the tandem mass spectrometry approach was developed and applied (Wilm et al., 1996). The tandem MS approach (also termed MS/MS or MS^n) enables one to fragment peptides and proteins in the gas-phase and, based on these fragment ions, the peptide sequence can be deduced (Wilm et al., 1996). Whole genome sequencing, development of new, faster and more sensitive MS instruments and miniaturizing liquid chromatography columns were start for mass-spectrometry based proteomics.

Besides mass spectrometry based proteomics, other methods exist to analyze the proteome. Some examples include cell imaging by light and electron microscopy, various electrophoresis and chromatography methods, array and chip experiments, and western blotting. Cell imaging allows one to localize and quantify proteins. Cryo electron tomography (Malmström et al., 2009) and fluorescence microscopy (Taniguchi et al., 2010) have been used to quantify proteins in single cells. Antibody based experiments such as enzyme-linked immunosorbent assay (ELISA) (Engvall and Perlmann, 1971) or more modern protein microarrays (reviewed by Berrade et al. (2011)) or western blotting (Renart et al., 1979) are useful for targeted detection and quantification of proteins. However, these techniques are limited in the number of proteins which can be identified and analyzed. To overcome this limitation there have been some major developments in protein microarrays to increase throughput (reviewed by Berrade et al. (2011)). Although the above mentioned methods are also important in proteome measurement, this dissertation is concentrated on mass spectrometry based proteomics.

2.1.2 *Proteome*

The term "proteome" was first used in 1994 by Marc Wilkins and was defined as "the entire **prote**in complement expressed by a gen**ome**" (Wilkins et al., 1996a). The human genome contains roughly 20,500 open reading frames (ORFs) which encode proteins (Clamp et al., 2007). However, it is believed that more than 2 million different proteins are expressed in different cells within the human body (Kelleher, 2012). This large diversity of proteins is driven mainly by alternative splicing (reviewed by Nilsen and Graveley (2010)) and a variety of post-translational modifications (PTMs) that further influence protein conformation and function (reviewed by Walsh et al. (2005)). Microbial proteomes are less complex, having less PTMs, and a smaller number of ORFs (*E. coli* 4,333 ORFs (Riley et al., 2006) and *Mycoplasma pneumoniae* has 690 ORFs (Maier et al., 2011)), making them ideal subjects to establish new molecular biology methods (*e.g.*, (O'Farrell, 1975; Hörth et al., 2006; Taniguchi et al., 2010)).

Proteins are built up from 20 amino acids with different side chains that define the chemical properties of amino acids. Covalent peptide bonds link amino acids to long polypeptide chains, which are known as primary structure of protein. Proteins fold into secondary structures (α-helices and β-sheets), which in turn fold into tertiary structure (three-dimensional organization). There is a large diversity of proteins: while there are only 20 amino acids, they can be combined into 20^n different polypeptide chains n amino acids long. However, not all of them are possible to exist due to unstable conformation (Alberts et al., 2002). After the proteins are translated they may undergo covalent modifications, called post-translational modifications (PTMs) (Walsh et al., 2005), to date over 200 of such modifications have been detected, the most important being phosphorylation, acetylation, methylation, glycosylation, disulfide bond formation, sulfation, hydroxylation, ubiquitination, carboxylation and acetylation of the N-terminal acid (Hoffmann and Stroobant, 2007). Proteins are the most important functional units of cells; they catalyze biochemical reactions, act as messengers and transporters and have also defense and structural roles. It is therefore understandable that there is a large interest in studying proteins and proteomes (reviews of Mallick and Kuster (2010); Walther and Mann (2010); Lamond et al. (2012)).

2.1.3 *Protein and peptide separation*

The goal of proteomics to analyze "the entire **prot**ein complement expressed by a gen**ome**" challenges scientists with complex mixtures of proteins with a wide dynamic range (*i. e.*, the concentration difference between the most and least abundant peptides). The complexity of the proteome can be reduced by protein and/or peptide fractionation. The first very powerful separating method applied on proteins was 2D gel electrophoresis (O'Farrell, 1975). Whilst having great resolving power, it is also very laborious and time consuming and therefore not suitable for high throughput proteomic workflows. Components of 2D gel electrophoresis, isolectric focusing (IEF) and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), are also used separately for reducing sample complexity prior liquid chromatography mass spectrometry (LC-MS).

IEF is based on the movement of a protein or a peptide in pH gradient and electric field towards its isoelectric point (Righetti, 1983). IEF as first dimension of protein or peptide fractionating method can be performed using immobilized pH gradient (IPG) gels (Cargile et al., 2004). Another approach is fractionation in the liquid phase, this has been referred to as off-gel method (Hörth et al., 2006; Hubner et al., 2008; Tran and Doucette, 2008b). Off-gel fractionation is advantageous over fractionation in the IPG gels due to better recovery of focused peptides or proteins, as no extraction from the gel needs to be performed (Hörth et al., 2006). Although isoelectric focusing of proteins can suffer from low recovery of alkaline and hydrophobic proteins, if performed on the peptide scale, at least some peptides from these problematic proteins can be typically detected (Hörth et al., 2006).

SDS-PAGE uniformly separates charged denatured proteins in a matrix made of cross-linked acrylamide that yields a porous network (Laemmli, 1970). Molecules move in an electric field towards the anode at different rates based on their size. SDS-PAGE can be used

also as a gel-free system where proteins are constantly eluted from the gel column and collected in the solution phase (Tran and Doucette, 2008a; Lee et al., 2009). Two dimensional liquid electrophoresis has been combined from off-gel IEF and gel-free SDS-PAGE to separate intact proteins (Tran et al., 2011). Gel-free systems are advantageous over in-gel systems due to the lack of protein recovery problems from the gel (Tran and Doucette, 2008a).

A common separation method for proteins and peptides is liquid chromatography (LC), where a liquid is used as the mobile phase and a porous solid as the stationary phase. The most often applied technique is reverse phase (RP) chromatography, where separation takes place due to an increase of organic solvent in the mobile phase, which detaches proteins or peptides from carbon chains in the stationary phase according to the hydrophobicity of the analytes. Shorter carbon chains (C_4, C_5, C8) are less retentive, and therefore used for intact protein separations (Capriotti et al., 2011), while longer chains (C18) are used for peptide separation. Size exclusion chromatography (SEC) has not seen widespread use in proteomics, because it offers relatively low peak capacity (Tran and Doucette, 2008a). However, it has been used to isolate large tryptic peptides from trypsin digested peptide pools for further digestion (Tran et al., 2011). A combination of strong anion exchange (SAX) or strong cation exchange (SCX) chromatography and RP chromatography are often combined and even automated to perform two dimensional separation, where molecules are separated first by charge and then by hydrophobicity (Washburn et al., 2001; Wagner et al., 2003).

Because each additional fractionation increases the required material and measurement time, one dimensional analysis with high proteome coverage would be preferable. Therefore, very long reverse phase columns (up to 50 cm) have been used recently to obtain deep coverage of proteomes (Thakur et al., 2011; Nagaraj et al., 2012; Cristobal et al., 2012). Long columns are packed with smaller particles to improve resolution, however, in order to reduce the back-pressure in regular LC, flow rates have to be reduced and columns should be heated (Thakur et al., 2011). Traditional high performance liquid chromatography (HPLC) has been improved with introducing ultra high performance liquid chromatography (UHPLC) (Plumb et al., 2004), where high pressure pumps are integrated in order to use less than 2 µm particles which produce significant increase in peak resolution, sensitivity and analysis speed (Plumb et al., 2004).

2.2 MASS SPECTROMETRY BASED PROTEOMICS

Today mass spectrometry based proteomics approaches allow one to identify and quantify thousands of proteins from complex samples in less than a day of acquisition time (Thakur et al., 2011). A common workflow for mass spectrometry based proteomics is presented on Figure 1. Briefly, proteins of interest are extracted from the sample and the complexity of sample is reduced by any of the protein and/or peptide separation described in the above section along with other methods. However, while potentially increasing proteome coverage, any form of fractionation requires more starting material and increases analysis time. These extra resources are not always acceptable for high throughput analysis of proteomes. Therefore, many of the bottom-up proteomics methods today are "shotgun" methods, which skip the pre-fractionation step and start by en-

zymatic cleavage of the proteins into peptides, most commonly with trypsin (Thakur et al., 2011). Separated peptides are ionized and entered into the mass analyzer where the mass to charge ratio (m/z) is measured and in case of a hybrid instrument, ions are fragmented for sequence determination.

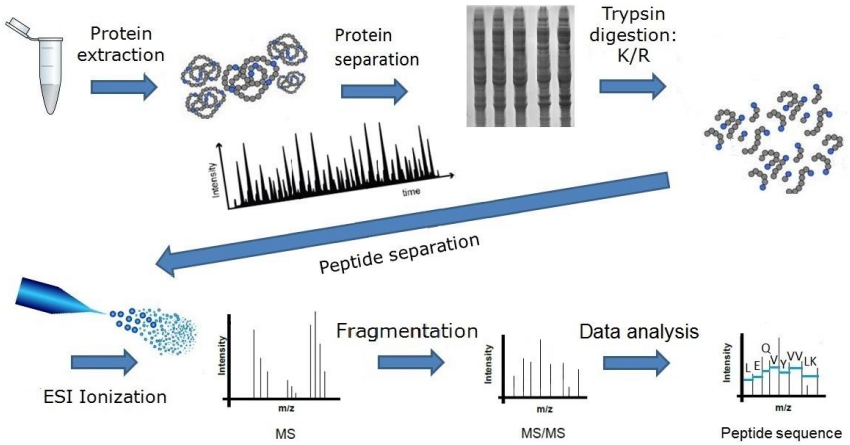


Figure 1 – Common workflow of bottom-up mass spectrometry based proteomics experiments.

2.2.1 *Mass spectrometers*

In general, a mass spectrometer consists of three basic components: an ion source that ionizes analytes, a mass analyser that measures the m/z value, and a detector that registers the m/z values.

Development of soft ionization techniques electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) enabled to analyze large, polar, and thermally labile biomolecules, such as proteins and peptides that previously were not possible to ionize (Fenn et al., 1989; Tanaka et al., 1988; Karas and Hillenkamp, 1988). Soft ionization refers to the ability to ionize and volatilize thermally labile compounds, such as peptides and proteins, without inducing any fragmentation (Mallick and Kuster, 2010). Other soft ionization methods less frequently applied in proteomics are atmospheric pressure chemical ionization (APCI) and fast atom bombardment (FAB) (Hoffmann and Stroobant, 2007). In addition to to soft ionization methods, also hard ionization methods exist such as electron impact, field ionization, *etc.* Hard ionization is usually applied to small, volatile molecules (Hoffmann and Stroobant, 2007). However, for proteomics research, the most often applied ion sources are MALDI and ESI.

For MALDI, the sample is first mixed and co-crystallized with a matrix, usually a UV-light adsorbing organic compound. In the ion source UV laser pulse is used to irradiate

the matrix, which leads to sublimation of the matrix and sample molecules into the gas phase. The analyte molecules are ionized by receiving protons from the ionized matrix molecules (Mallick and Kuster, 2010). MALDI generated ions are singly charged and thus are favorable for analyzing intact proteins (Yates et al., 2009). Because MALDI is not directly compatible with LC, it is mostly used to analyze pure molecules or simple mixtures, which do not require any additional separation prior to mass spectrometry (MS). However, HPLC separated peaks can be collected and spotted into a matrix on the MALDI target by robotics and analyzed by MALDI (Mirgorodskaya et al., 2005).

ESI has an advantages over MALDI due to the fact that it produces ions from solution and therefore is compatible with LC. ESI is performed by applying a high (1-6 kV) voltage to the capillary carrying the sample in liquid flow, and results in an electrically charged spray of droplets (Yates et al., 2009). There are two theories regarding how ionization works in ESI: (I) after all the solvent is evaporated, excess charges remain on the analytes (Iribarne and Thomson, 1976), and (II) analytes are accumulated at the surface of the droplet and are extracted and ionized by field desorption (Dole et al., 1968). In ESI, peptides are mainly multiply charged, as roughly each basic site in the peptide gets protonated (Mallick and Kuster, 2010). One of the most important developments in proteomics was the invention of nano-ESI (Wilm et al., 1996) compatible with nanoflow HPLC (Shen et al., 2002). A nanoflow HPLC system is characterized by very small column diameters (*e.g.*, 75 µm) and operates at low flow rates (usually in the range of 50-800 $\mathrm{nL{\cdot}min^{-1}}$). This is advantageous because peptides can be ionized much more effectively if they are concentrated into small droplets (Gibson et al., 2009). Furthermore, MS is a concentration dependent instrument and if the same amount of analyte is concentrated to smaller volume, MS signal will be amplified (Shen et al., 2002). A nano-ESI emitter can be integrated with a capillary column to minimize post-column dead volume which affects the separation quality (Xie et al., 2006). A disadvantage of the capillary merged emitter is that if the emitter fails, it renders the column unusable (Shen et al., 2002).

There are currently five types of mass analyzers used in proteomics: time-of-flight (TOF), quadrupole, linear ion-trap, Fourier transform ion cyclotron resonance (FT-ICR) and Orbitrap mass analyzers (review of Mallick and Kuster (2010)). In order to perform fragmentation of molecules, tandem mass spectrometers are used where different mass analyzers are connected, *i. e.*, hybrid instruments.

In a time-of-flight (TOF) mass analyzer, the m/z value of an ion is detected by measuring the time it takes for an ion to travel over a fixed distance inside the high vacuum of the mass analyzer (review of Mallick and Kuster (2010)). TOF MS is often used as a single analyzer in combination with a MALDI or ESI ionization source (Eidhammer et al., 2008).

A quadrupole mass analyzer consists of four parallel rods of electrodes. A strong electric field between the electrodes ensures that only ions of defined mass can pass through the electrodes (Eidhammer et al., 2008). An oscillating electrostatic field forces ions to follow a spiral trajectory through the quadrupole rods, with the radius of the ion spiral depending on the m/z value. Ions with a specific m/z value can be trapped with a specific oscillating field between the electrodes for fragmentation, while the majority of ions are discarded (Eidhammer et al., 2008; Schuchardt and Sickmann, 2007).

Quadrupoles are used most commonly as part of a hybrid instrument, *e.g.*, for accumulating, isolating and fragmenting the ions emitted from the ion source on the way to an-

other mass analyzer (May et al., 2011). A combination of three quadrupoles forms a triple quadrupole (QqQ) system, where the first quadrupole is used as a mass filter, the second as a collision cell, and the third as the detector. QqQs systems are relatively slow, low accuracy (100 ppm) and low resolution (up to 2,000 full width at half maximum (FWHM)) instruments, however they make up for these drawbacks with their excellent sensitivity (down to attomole level) and dynamic range (up to six orders of magnitude) (Yates et al., 2009). For Q-TOF two quadrupoles are combined with a TOF analyser. In the MS mode, the quadrupoles act as an ion guide to the TOF analyzer where mass analysis takes place. In MS/MS mode, the precursor ions are selected in the first quadrupole and undergo fragmentation in the second quadrupole. The product ions are analyzed in the TOF device (Domon and Aebersold, 2006). Q-TOF instruments have a resolution up to $>$ 25,000 FWHM (Domon and Aebersold, 2010), a mass accuracy of 2-5 ppm, sensitivity up to attomole levels and a dynamic range of six orders of magnitude (Yates et al., 2009).

A linear ion trap consists of four parallel rods and functions similarly to triple quadrupole, however, ion selection, fragmentation, and mass analysis are all performed in a single device at different times (Schuchardt and Sickmann, 2007). If an ion trap is over filled, ions start to interact with each other and a significant loss of resolution and mass accuracy will follow due to this "space charge" phenomenon (Hager, 2002). To minimize this saturation effect, the amount of ions collected in the ion trap must be optimized. Another peculiarity of an ion trap is the low m/z cut-off as ions with m/z values of less than 30% of precursor m/z are not trapped (also called 1/3 rule) (Cunningham et al., 2006). The fast scanning rate, sensitivity, flexibility, robustness, and relative low cost are the advantages of ion trap mass analyzers (Schuchardt and Sickmann, 2007). Low mass resolution (2,000 FWHM) and low accuracy (100 ppm) are the main disadvantages (Domon and Aebersold, 2010). Hybrid instruments where a linear ion trap replaces the third quadrupole in a triple quadrupole system is termed Q-trap. These instruments have sensitivity down to the attomole level while still having the speed of a linear ion trap and resolution, accuracy, and dynamic range similar to QqQ (Yates et al., 2009).

Fourier transform ion cyclotron resonance (FT-ICR) MS was for a long time considered to be the most accurate and suitable instrument for intact protein analysis (Ahlf et al., 2012). The ions in FT-ICR MS are trapped in a cyclotron where the combination of electric and strong magnetic fields accelerates ions to high energy. A Fourier transformation is used to generate the m/z signal (Eidhammer et al., 2008). FT-ICR MS has the highest resolution (500,000 FWHM) and mass accuracy ($<$ 2 ppm) of any other currently available mass spectrometer (Schuchardt and Sickmann, 2007; Yates et al., 2009). FT-ICR has a femtomole sensitivity and dynamic range of four orders of magnitude (review of Yates et al. (2009)). However, as FT-ICR MS needs high magnetic fields it makes the technology very cost-intensive and poorly accessible (Schuchardt and Sickmann, 2007).

A major technological improvement in proteomics was the development of the Orbitrap mass detector (Olsen et al., 2005), which has made rapid, high-sensitivity mass spectrometry more affordable and more widely available (Vogel and Marcotte, 2012). The Orbitrap can be regarded as a highly modified ion trap (Schuchardt and Sickmann, 2007), it consists of two concentric electrodes and uses orbital trapping of ions in its electrostatic fields. The ions orbit around a central electrode and oscillate in the axial direction. No magnetic fields are involved and the m/z is obtained by a Fourier transformation of

the ion current (Olsen et al., 2005). When coupled to an ion trap, the hybrid instrument retains the advantages of both: four orders of dynamic range, high resolution (more than 100,000 FWHM), high mass accuracy (down to sub-ppm), and good speed and the sensitivity (at femtomole range) (Yates et al., 2009). Orbitrap mass analyzers equipped with a quadrupole mass filter in the front allows one to select precursor ion mass and perform targeted analysis — selected ion monitoring (SIM) and selected reaction monitoring (SRM), with a high resolution (up to 140,000 FWHM) (Michalski et al., 2011b; Gallien et al., 2012).

2.2.2 *Protein/peptide fragmentation and identification*

Ions are fragmented in a mass spectrometer in order to obtain information about their structure or sequence. Proteins and peptides are mostly fragmented by collision induced dissociation (CID), however also other fragmentation methods exist: electron capture dissociation (ECD) (Zubarev et al., 1998), electron transfer dissociation (ETD) (Syka et al., 2004), higher energy collisional dissociation (HCD) (Olsen et al., 2007). Despite the fragmentation method used, the general fragmentation pattern of peptides is the same: the fragments are labeled as a, b, c for N-terminally and x, y, z for C-terminally cleaved bonds (Figure 2) (Roepstorff and Fohlman, 1984; Biemann, 1988).

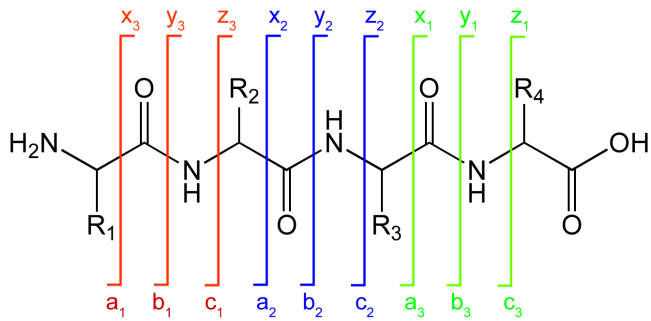


Figure 2 – Peptide backbone fragmentation pattern: N-terminal fragments a, b, c and C-terminal fragments x, y, z.

In CID, fragmentation is induced by collision of ions with residual gas in the tandem MS collision cell and in case of peptides, fragmentation mainly occurs through the cleavage of the peptide bond, producing y and b fragments (Steen and Mann, 2004). CID preferably breaks the weakest bonds, a clear disadvantage in post-translational modification studies, where labile modifications are fragmented first (Hoffmann and Stroobant, 2007). The same applies for large peptides and proteins where mostly N- and C-terminal regions are fragmented (Eidhammer et al., 2008; Hoffmann and Stroobant, 2007).

CID can also be applied in one of the high-pressure regions between the atmospheric pressure source and the mass spectrometer (van Dongen et al., 1999). This process is called "up-front CID" (van Dongen et al., 1999) and has also been referred as "in-source

CID" (Williams et al., 2003). This method has been applied to identify simple mixtures of tryptic peptides (Williams et al., 2003) and rapid top-down characterization of antibodies (Zhang and Shah, 2007; Ren et al., 2009). Because there is no precursor ion selection for fragmentation, observed fragments can result from all precursor ions, including from the solvent and background species, which may result in a noisy spectrum (Williams et al., 2003).

CID fragmentation in linear ion traps is highly efficient, however, when recorded in the same linear ion trap, the resulting MS/MS spectra have relatively low mass accuracy and resolution (Olsen et al., 2007). Furthermore, low-mass fragment ions are not trapped (Cunningham et al., 2006), making small N- and C-terminal fragment ions unobservable. Because CID fragmentation is preferably carried out through the lowest energy fragmentation pathways, this sometimes leads to uninformative spectra (Olsen et al., 2007). Pulsed Q collision induced fragmentation (PQD) (Schwartz et al., 2005) and higher energy collisional dissociation (HCD) (Olsen et al., 2007) were developed to overcome the limitations of CID in trapping instruments and their hybrids. Although covering low m/z range, PQD has been reported to be less effective for fragmentation than CID (Bantscheff et al., 2008). HCD is more effective than CID and produces better coverage of the low mass region and more y fragments. This technique is beneficial for the quantification of isobaric tag (TMT and iTRAQ) labelled peptides in the low mass region (Köcher et al., 2009; Pichler et al., 2011), and for better peptide coverage in order to perform *de novo* sequencing (Michalski et al., 2012b).

While CID, HCD, and PQD mostly produce b and y fragments, there exist complementary fragmentation methods that produce different fragments that increase the sequence coverage of peptides and proteins. Electron capture dissociation ECD occurs in FT-ICR MS when trapped multiple protonated ions are exposed to thermal electrons and fragment mostly to c and z ions, and occasionally also to a ions (Zubarev et al., 1998). However, because thermal electrons are not possible to trap in ion traps and triple quadrupoles, ECD is exclusively used only with FT-ICR MS (Syka et al., 2004). ETD (Syka et al., 2004) produces similar fragments to ECD and can be used in triple quadrupole systems, linear ion traps, and their hybrid instruments. In ETD, electrons are delivered to multiply charged molecules with anions which can be effectively trapped in triple quadrupoles and ion traps (Syka et al., 2004). ECD and ETD are preferred fragmentation methods for proteins because they can fragment molecules nearly everywhere along the backbone (Hoffmann and Stroobant, 2007) while CID causes fragmentation mainly at the terminal regions (Eidhammer et al., 2008). Both, ECD and ETD, are for the same reason also used to study modified peptides or proteins (Xu et al., 2011).

Although it is possible to identify proteins based on their proteolytic peptide masses — peptide mass fingerprinting (Yates III et al., 1993; Henzel et al., 1993; Pappin et al., 1993), information rich fragmentation spectra are helpful in many ways for peptide and protein identification. First, it is possible to perform *de novo* sequencing for unknown peptides by reading out peptide sequence based on the mass differences of fragments that correspond to individual amino acids (Ma and Johnson, 2011). Second, biological samples often contain complex mixtures of proteins where peptide mass fingerprinting is not capable of mapping the peptides back to proteins (Eng et al., 1994). Third, if proteins or peptides

contains many modifications or unknown modifications, it is impossible to identify them based only on the exact mass (Lanucara and Eyers, 2012).

Tandem MS spectra can be identified by *de novo* sequencing, where the peptide sequence is read from MS/MS spectra by interpreting fragment ions, either manually, or with the help of special software (PEAKS, PepNovo) (review of Nesvizhskii et al. (2007)). *De novo* sequencing is advantageous if the organism is unknown or the protein has some unknown modifications (review of Nesvizhskii et al. (2007)). However, the high-throughput interpretation of tandem MS spectra is performed by search engines which compare precursor masses and their fragments with peptides from *in silico* digested proteome databases (Nesvizhskii et al., 2007). The candidate peptides are restricted to specified criteria such as mass tolerance, digestion enzyme, and PTMs allowed, and the search engine will result in a list of peptides that match with the measured MS/MS spectra (Nesvizhskii et al., 2007). Search engines are mainly applying probability based algorithms such as Mascot (Perkins et al., 1999) and Andromeda (Cox et al., 2011), or heuristic algorithms like SEQUEST (Eng et al., 1994) and XTandem (Craig and Beavis, 2004). Probability based algorithms calculate a probability that the observed number of matches between the calculated and measured fragment masses could have occurred by chance (Cox et al., 2011). Heuristic algorithms correlate acquired MS/MS spectra with theoretical spectrum and counts the number of peaks in common (Eng et al., 1994).

In large-scale proteomics studies, database searching remains the most frequently used peptide identification method (Nesvizhskii et al., 2007). However, there is also the possibility of identification based on sequence tags. Sequence tags combine small *de novo* identified fragments and database searchs to identify peptides (Mann and Wilm, 1994) and proteins (Mørtz et al., 1996). This method starts by identifying short partial sequences from MS/MS spectra, followed by a database search (Nesvizhskii et al., 2007).

In order to validate the identified database identification, false discovery rate (FDR) is calculated, which is identified as the proportion of incorrect identification among all identifications (Nesvizhskii et al., 2007). FDR estimation can be carried out by a targeted-decoy strategy, where MS/MS spectra are searched against a target database of protein sequences supplemented with the reversed or randomized sequences of the same database (Mascot, Andromeda) (Perkins et al., 1999; Cox et al., 2011). A FDR threshold will be applied to filter the data to sufficient reliability (Cox et al., 2011). Different strategies may be applied to search against protein sequence databases and to calculate the probability of each peptide being correctly assigned. It is also possible to estimate the FDR from this probability (peptideProphet and ProteinProphet) (Nesvizhskii et al., 2003). In addition to FDR, also false positive rate (FPR) is sometimes used. It is defined as the probability that a randomly matched spectrum is correctly matched (Nesvizhskii et al., 2007).

2.2.3 *Top-down and bottom-up proteomics*

In mass spectrometry-based proteomics, three approaches exist how to analyze proteins: the top-down approach that measures intact proteins; the middle-down approach that measures long polypeptides, and the bottom-up approach that cuts proteins into peptides and identifies proteins based on the peptides present.

Top-down proteomics analyses intact protein masses and their fragments. Fragmentation is obtained most preferably by ECD and ETD gas-phase dissociation (Yates et al., 2009), although on Orbitrap platforms, HCD has also been used successfully (Ahlf et al., 2012; Michalski et al., 2012a). Top-down proteomics identifies proteins using sequence tags (Mørtz et al., 1996) or based on accurate masses (Fenn et al., 1989) or more often combines both (Durbin et al., 2010). In order to successfully deconvolute multiply charged ions in ESI spectra to singly charged species, the MS must be able to resolve isotope distributions and ion charge states; therefore, high resolution instruments like FT-ICR MS and Orbitrap are mainly used (Hoffmann and Stroobant, 2007; Durbin et al., 2010).

By top-down methods combinations of protein PTMs are easier to characterize (Lanucara and Eyers, 2012) and better sequence coverage of the protein is obtained allowing to identify protein isoforms (Tran et al., 2011). It is also more precise to directly quantify intact proteins than to calculate protein concentration based on the peptides originating from the protein (Du et al., 2005; Waanders et al., 2007). Top-down proteomics measuring intact proteins is gaining more attention as the protein separation methods improve (Tran et al., 2011; Ahlf et al., 2012) and software develops (Durbin et al., 2010). Recently top-down proteomics approach applying Orbitrap mass analyzer and gel-free electrophoresis identified 690 unique proteins and over 2000 protein isoforms (proteoforms) with intact masses < 50 kDa (Ahlf et al., 2012).

Although it seems to be more reasonable to analyze proteins directly rather than cut them into peptides and then attempt to identify them from a complex mixture, top-down proteomics approaches are still too cumbersome for routine use. For example, synchronizing protein on-line separation with the MS time scale is challenging because intact proteins require more time for ionization and fragmentation (Lanucara and Eyers, 2012) and most often off-line LC or gel-free IEF or/and SDS-PAGE fractions are analyzed (Tran et al., 2011; Ahlf et al., 2012; Lanucara and Eyers, 2012). Furthermore, top-down analysis is limited to high mass-accuracy and high resolution instruments such as FT-ICR MS (Tran et al., 2011; Durbin et al., 2010) and Orbitrap (Ahlf et al., 2012) in order to interpret the isotopic patterns and calculate the exact mass of a protein.

Bottom-up approachs digest proteins to peptides by enzymes which are sequence specific, meaning their cleavage sites are known. Trypsin is often used because it produces peptides that have a suitable average length for MS analysis and which also fragment well (Steen and Mann, 2004). The specificity of trypsin is to cut C-terminal ends to lysine (Lys) and arginine (Arg) residues, producing mainly peptides which are 6-25 amino acids long and have basic residue that are easily protonated (Steen and Mann, 2004). Bottom-up proteomics is a preferred method because intact protein separation prior to MS and MS analysis is considered to be too difficult.

Digesting an entire proteome into peptides increases its complexity and therefore more effort is required to fractionate it down to individual species. Therefore bottom-up proteomics has concentrated on exhaustive fractionation steps (Washburn et al., 2001; de Godoy et al., 2008). Currently, one of the most promising bottom-up approaches in high throughput proteomics is a shotgun approach based on only long reverse phase chromatography separations of tryptic peptides (Thakur et al., 2011; Nagaraj et al., 2012; Cristobal et al., 2012).

Middle-down proteomics combines top-down and bottom-up strategies by analyzing long polypeptides (3,000-20,000 Da), obtained by digesting with enzymes that cleave between less common sites (Cannon et al., 2010; Wu et al., 2012). Due to the long sequences analyzed, middle-down proteomics results in high confidence identifications and good sequence coverage (Cannon et al., 2010).

2.2.4 *Mass spectrometry acquisition modes*

There are several acquisition modes which can be used for different proteomics experiments: data dependent analysis (DDA), data independent analysis (DIA) (Venable et al., 2004) and selected reaction monitoring (SRM; plural multiple reaction monitoring (MRM)) (Lange et al., 2008).

In DDA experiments, a full MS scan of all the precursor ions is followed by selecting a number of precursor ions (typically 5-10, however it has expanded with current instruments to even up to 25 ions (Thakur et al., 2011)) for MS/MS fragmentation based on their abundance, and then dynamically excluded from fragmentation for a certain time, usually 30-90 seconds (Hoehenwarter and Wienkoop, 2010). The DDA approach has been used extensively so far for most of the proteomics studies (Ishihama et al., 2008; de Godoy et al., 2008; Schwanhäusser et al., 2011; Nagaraj et al., 2012) due to the speed limitation of MS. However, as peptides can co-elute from the column and are selected for fragmentation on the basis of their abundance, this might affect the identification of low abundant peptides. It has been demonstrated that if a complex cell lysate is analyzed by a standard LC run, only small fraction of detectable peptides are fragmented in DDA (Michalski et al., 2011a). This is mainly due to the high complexity of the sample, and if analysis speed would increased by 10 times, most of the peptides would be identified (Michalski et al., 2011a).

The data independent analysis (DIA) strategy has been developed to complement the DDA method for proteomic analysis (Venable et al., 2004). Instead of a serial selection of precursor ions for data dependent fragmentation, the DIA approach fragments a group of co-eluting precursor ions at each given time, enabling a more unbiased detection of all LC-eluted peptides compared to the DDA method (Venable et al., 2004; Ramos et al., 2006; Purvine et al., 2003). The DIA strategy is currently commercially available on three MS platforms: MS^E (Silva et al., 2006b) on a quadrupole time-of-flight (Q-TOF) instrument (Waters); "all-ion fragmentation" (AIF) (Geiger et al., 2010a) on an Orbitrap mass analyser (Thermo Fisher Scientific) and SWATH (Gillet et al., 2012), on a triple quadrupole-TOF instrument (AB Sciex). In Q-TOF alternation between low and elevated energy in the quadrupole is used; Orbitrap alternates MS scans between full scans with HCD fragmentation applied in the collision cell (Silva et al., 2006b; Geiger et al., 2010a). For SWATH analysis the first quadrupole passes small m/z ranges to a collision cell where all ions are fragmented and fragments are analyzed in TOF (Gillet et al., 2012), those swaths are done consecutive to cover large m/z range.

The major challenge of DIA acquisition is that the direct relationship between the precursor and its fragments is lost, and the fragment spectra can be sometimes very difficult to interpret. In most studies this problem has been alleviated by making use of the fact

that precursors and fragments must "co-elute" in time (Geiger et al., 2010a) *i. e.*, chromatographic retention time of the precursor and fragment masses are very similar or in case SWATH method the small m/z windows will give also additional confidence in precursor-fragment pairs (Gillet et al., 2012).

SRM analysis is a targeted method well suited for detecting and quantifying specific proteins (Picotti et al., 2009; Costenoble et al., 2011). SRM analysis is mainly used in QqQ, where the first quadrupole acts as a mass filter, allowing through only selected m/z range which will be fragmented in the second quadrupole (Lange et al., 2008). The third quadrupole then acts as a mass filter of resulting fragment ions in a similar way to the first quadrupole, allowing through only fragment ions of a particular, selected mass (Lange et al., 2008). Triple quadrupole ion trap has demonstrated SRM analysis sensitivity with combination to off-gel IEF by detecting proteins expressed at a single-digit number of copies/cell (Picotti et al., 2009). Recently, it has become possible to perform SRM with a quadrupole Orbitrap hybrid instrument as well (Gallien et al., 2012). SRM is a preferred method for targeted quantification because it is able to consistently record the intensities of predefined target fragment ions across the analysis (Gillet et al., 2012). However, SRM is limited to the measurements of few thousands of transmissions (Costenoble et al., 2011) and therefore not compatible with large scale proteome analysis (Gillet et al., 2012).

Trapping instruments can use similar analysis — selected ion monitoring (SIM), which scans only a very narrow m/z range and ions outside of this range are not scanned. SIM has been used in an ion trap hybrid instruments for absolute quantification down to the attomole level (Hanke et al., 2008) by using DDA with an inclusion list containing the masses of the expected peptides of the standard protein.

2.3 QUANTITATIVE MASS SPECTROMETRY-BASED PROTEOMICS

Mass spectrometry-based proteomics turned quantitative shortly after its birth (Wilm et al., 1996) as it was realized that protein identification only provides only very limited information (Mann, 1999; Ong and Mann, 2005). While a number of other methods exist, this overview will concentrate on bottom-up gel-free quantification approaches.

There are two main approaches used in quantitative mass spectrometry based proteomics: stable isotope labeling and label-free quantification. Stable isotope labeling is more accurate quantification method, while label-free quantification allows one to perform comparisons across many samples, and there is no need for expensive labeled substances. Mass spectrometry-based quantitative proteomics can also be classified into relative and absolute quantification. In this dissertation, relative quantification refers to comparison of the measure of the same protein present in different samples, while absolute quantification refers to the amount (*e.g.*, fmol, molecules, µg) of a protein in the sample regardless of the measurement method applied (label-free or stable isotope based).

2.3.1 *Stable isotope labelling*

Quantification with stable isotopes is based on the mass difference between labeled and unlabeled ions in MS analysis. After mixing samples, the intensity ratio between the iso-

tope variants reflects the difference between their abundances. If protein concentrations for one of the samples are known, also accurate absolute quantification could be carried out.

Stable isotopes can be incorporated into proteins by *in vivo* or *in vitro* methods. In *in vivo* techniques, isotope enriched compounds are incorporated from the growth media into all of the proteins in the organism under study (Gouw et al., 2010). *In vitro* techniques use chemical reactions for incorporation of the stable isotopic tags onto selective sites on peptides of proteins (Yan and Chen, 2005).

The most precise method for absolute protein quantification is spiking a sample with known amounts of isotopically labeled standard peptides (Gerber et al., 2003), concatenated peptides (Pratt et al., 2006), or proteins (Brun et al., 2007; Hanke et al., 2008; Matic et al., 2011). To reduce interference from background ions, quantification of isotope labeled standards can be performed on specific fragments of the peptide using SRM (Lange et al., 2008). While the advantages of using isotopically labeled standards have been reviewed (Brun et al., 2009; Pan et al., 2009), there are also many disadvantages in terms of synthesizing, storing and handling the standards (Mirzaei et al., 2008; Hanke et al., 2008), accurate quantification of standards (Hanke et al., 2008), cost (Ludwig et al., 2012) and difficulties to have full coverage of complex proteome (Brownridge et al., 2011).

In order to overcome the above mentioned limitations of stable isotope standards one could use enzymatic or chemical labeling. Enzymatic labeling of proteins by digesting in $H_2{}^{18}O$ incorporates two labeled oxygen atoms into the enzyme-cleaved peptides, which allows to distinguish heavy species from light species in MS by 4 Da mass difference between them (Schnölzer et al., 1996). Although it is very simple method the small (4 Da) mass difference can overlap natural isotopes and also incomplete labeling has been reported (Fenselau and Yao, 2009). Chemical labeling can be performed on the protein level by isotope-coded affinity tag (ICAT) (Gygi et al., 1999a), iTRAQ (Wiese et al., 2007), *etc.*; or on the peptide level by TMT (Thompson et al., 2003), iTRAQ (Ross et al., 2004), dimethyl labeling (Hsu et al., 2003; Boersema et al., 2009), *etc.* While mass differences introduced by enzymatic labeling, ICAT and dimethyl labeling are detected on MS1 level; isobaric chemical labeling methods iTRAQ and TMT labels are distinguished on MS/MS level. Isobaric labels have the same mass but different fragments in low m/z range of MS/MS spectra that are used for quantification. As for ion traps the recovery of fragment ions below 30% of the precursor ion mass is very poor (Cunningham et al., 2006), iTRAQ and TMT are challenging methods for ion trap hybrid instruments (Bantscheff et al., 2008; Köcher et al., 2009; Pichler et al., 2011). However, the great advantage of iTRAQ and TMT is the possibility to perform multiplexed quantification of up to eight samples at the same time (Choe et al., 2007). This saves instrument time and simplifies experimental design (Li et al., 2011).

While chemical and enzymatic labeling kits are commercially available and generally easy to use with well-established protocols, they tend to be expensive due to costly chemicals. Also, because mixing of the samples takes place after the digestion, quantification accuracy suffers (Ong and Mann, 2005). Additionally, there are some limitations for MS, as small m/z differences require high resolution instruments in order to resolve the isotopic patterns (Fenselau and Yao, 2009).

Considering the proteomics workflow in Figure 1, and all of the steps taken in order to analyse proteins and peptides in MS, it is obvious that the best time to introduce an internal standard would be before protein extraction from the sample *i. e.* by metabolically incorporating them into the living organism or cells. This approach produces the lowest bias to quantification (review of Gouw et al. (2010)). Unlike other labeling technologies, samples to be analyzed are combined before protein extraction and digestion, thus removing the main source of uncontrolled sample variability (review of Bantscheff et al. (2007)). Therefore, metabolic labeling is suitable for samples that need to undergo extensive preparation steps at the protein or peptide level, such as fractionation and enrichment, which may introduce a significant amount of error (Li et al., 2011).

Metabolic labeling for quantifying the proteome was first applied in yeast (Oda et al., 1999), it involved comparison of two states by growing the yeast on media enriched with ^{15}N in one state and on media containing the naturally abundant isotope ^{14}N in the other state. The ratios of $^{14}N/^{15}N$ containing proteins from the two conditions were measured by MS and changes in protein expression levels were determined. ^{15}N labeling has been applied in top-down (Du et al., 2005) as well as bottom up protein quantification studies (Hendrickson et al., 2006; Palmblad et al., 2007; Nelson et al., 2007; Li et al., 2011).

A disadvantage of ^{15}N-labeling is the fact that the mass difference between the unlabeled and ^{15}N-labeled peptides is unknown during mass spectrometric analysis and becomes apparent only after peptides are identified, because each peptide incorporates a different number of nitrogen atoms depending on the length of the peptide and the number of amino acids that contain nitrogen atoms on side chains (Gouw et al., 2010). Therefore, stable isotope labeled amino acids have been used to alter the mass of proteins (Ong et al., 2002). The use of labeled amino acids can be carefully selected and even multiplexed (Blagoev et al., 2004). In the stable isotope labeling by amino acids in cell culture (SILAC) method cells are labeled through the incorporation of stable labeled heavy essential amino acids, typically $^{13}C_6{}^{15}N_2$-lysine and $^{13}C_6{}^{15}N_4$-arginine, which are best suited for trypsin digestion (Ong et al., 2002). SILAC can be used for triple quantification (Blagoev et al., 2004) and this has been utilized in pulsed SILAC experiments (Schwanhäusser et al., 2009) to measure translation rates and protein turnover (Schwanhäusser et al., 2011; Boisvert et al., 2012). Furthermore, SILAC is not limited to only three labels: four different heavy stable isotopic forms of arginine ($^{13}C_4$, $^{13}C_6$, $^{13}C_6{}^{15}N_4$, $^{13}C_6{}^{15}N_4{}^{2}H_7$) combined with light arginine have been used for 5plex quantification (Molina et al., 2008). Also, an absolute SILAC approach has been proposed (Hanke et al., 2008) where SILAC-labeled proteins are produced *in vivo* or *in vitro* (Matic et al., 2011) and added to samples for absolute quantification.

One major consideration in choosing a metabolic labeling scheme is whether the cells studied can metabolically incorporate the labeled precursors or not (Ong and Mann, 2007). For example, labeling cells with amino acids, as it is done in the SILAC approach, is possible only if cells are auxotrophic for the labeled amino acid(s) and not synthesized by the cells themselves (Hanke et al., 2008). There are no special requirements to the metabolism of cells if salts labeled with ^{15}N are used for metabolic labeling — in this case the label is incorporated into all amino acids even if the cells are capable of synthesizing all amino acids in the biomass (review of Gouw et al. (2010)). Conversion of labeled arginine to proline can occur in certain strains or cell types (Ong and Mann, 2007). Although

such conversion is undesirable for SILAC quantification, it does not affect ^{15}N labeling because all amino acids are labeled (review of Gouw et al. (2010)). Metabolic labeling was considered to be limited to use only in cell cultures, this has been challenged by development of SILAC mouse (Krüger et al., 2008), worm (Larance et al., 2011), fly (Sury et al., 2010). Super-SILAC, which mixes multiple SILAC-labeled cell lines, is used as internal standard for human tumor tissue quantification (Geiger et al., 2010a). Super-SILAC has made metabolic labeling possible also for human tissues and brought metabolic labeling towards clinical relevance (Geiger et al., 2010a).

2.3.2 *Label-free quantification*

Label-free quantification is an alternative quantification method which compares separately prepared and analyzed LC-MS/MS runs. It is widely used because it skips the laborious and costly process of introducing stable isotopes and is applicable to samples from any source (Li et al., 2011). The most simple label-free quantification technique is spectral counting, which is based on the observation that the more abundant the protein is, the more peptides can be identified from a protein (Washburn et al., 2001). Because larger proteins produce more peptides and therefore also more MS/MS events, spectral counting methods must take into account the size of proteins. This has been implemented in several spectral counting methods: protein abundance index (PAI) (Rappsilber et al., 2002), exponentially modified protein abundance index (emPAI) (Ishihama et al., 2005), normalized spectral abundance factor (NSAF) (Zybailov et al., 2006), absolute protein expression (APEX) (Lu et al., 2007), *etc.* The normalized spectral index (SI_N) combines peptide count, spectral count and fragment ion intensity (Griffin et al., 2010).

The emPAI method is based on comparing the number of experimentally observed peptides and calculated number of observable peptides (Ishihama et al., 2005). emPAI is an improvement of PAI (Rappsilber et al., 2002), which is defined as follows:

$$\mathrm{PAI} = \frac{N_{obsd}}{N},$$

where N_{obsd} is the number of experimentally observed peptides per protein and N is the the number of theoretically observable peptides per protein (Rappsilber et al., 2002). The emPAI is defined as follows (Ishihama et al., 2005):

$$\mathrm{emPAI} = 10^{\mathrm{PAI}} - 1.$$

Although such a method of concentration determination may not be very precise, the accuracy of concentration measurements using emPAI values were demonstrated to lie within the same error range or even better than protein concentration measurements based on staining methods such as the Bradford assay (Ishihama et al., 2005). emPAI has been used to measure protein abundance for more than 1,000 *E. coli* proteins with good agreement with 40 enzymes of known amount (Ishihama et al., 2008). Because emPAI is implemented into the Mascot database search platform, it is possible to apply this approach to previously measured or published datasets to add quantitative information without any additional steps.

While emPAI employs unique precursor ions (Ishihama et al., 2005), another spectral counting method APEX uses the total number of MS/MS scans observed for peptides from a protein (Lu et al., 2007). APEX is defined as:

$$\mathrm{APEX}_i = \frac{n_i \times p_i}{O_i \times \sum_{k=1}^{\#\text{ observed proteins}} \frac{n_k \times p_k}{O_k}} \times C,$$

where C is the total concentration of protein molecules in the sample, n_i is the total number of MS/MS scans observed from peptides of protein i through the course of the experiment, p_i is the probability of correctly identifying the protein i, O_i is an estimate of the number of expected unique peptides observed for protein i (Lu et al., 2007).

The critical correction factor in APEX is O_i, which is calculated for each protein separately by training a classification algorithm to predict the observed tryptic peptides from a given protein based upon peptide lengths and amino acid compositions (Lu et al., 2007). Lu et al. (2007) reported O_i to improve estimation of protein abundance by up to ~30%. APEX successfully determined the abundance of 10 proteins that were spiked in a yeast cell extract with known amounts. In addition, the absolute protein abundance of yeast and *E. coli* proteomes analyzed by APEX correlate well with the measurements by western blotting and flow cytometry (Lu et al., 2007). Calculation of APEX values can be carried out by freely available software: APEX Quantitative Proteomics Tool (Braisted et al., 2008).

Spectral counting approaches have been reported to be not particularly sensitive to small changes in abundance (Hendrickson et al., 2006) and less reproducible than labeling based approaches (Li et al., 2011). It has also been demonstrated that at higher concentrations (> 100 fmol on column) and in complex samples, spectral counting methods suffer from saturation effects (Ishihama et al., 2008; Grossmann et al., 2010). Optimization of dynamic exclusion settings in DDA mode can increase the reproducibility of spectral counting and the quantification of low abundant proteins (Hoehenwarter and Wienkoop, 2010).

Peak intensity based quantification is an alternative to spectral counting label-free quantification techniques where the mass spectrometric areas or intensities of peptides in each of the experiments are measured (Chelius and Bondarenko, 2002). There are several peak intensity based methods: high and low collision energy switching (MS^E) used in DIA mode (Silva et al., 2006a) and T3PQ used in DDA mode (Malmström et al., 2009; Grossmann et al., 2010) are based on the intensities of three most intense tryptic peptides from a protein. Intensity based absolute quantification (iBAQ) takes into account the sum of all identified peptide intensities and divides it by the number of theoretically observable peptides (Schwanhäusser et al., 2011). iBAQ absolute quantification is based on standard curve combined from 48 accurately quantified human proteins (UPS2, Sigma-Aldrich); dynamic range of concentrations spanning six orders of magnitude. Linear regression is used to fit iBAQ intensities to absolute amounts of standard proteins (UPS2 standard) amounts (Schwanhäusser et al., 2011):

$$\mathrm{iBAQ} = \frac{\sum_{n=1}^{N} I_n}{N_{obs}} \times a + b,$$

where N is the number of unique precursors, and N_{obs} is the number of theoretical peptides by *in silico* protein digestion, I_n is the maximum peak intensity for a peptide, a and b are the slope and intercept determined by standard curve of spike in proteins (UPS2) (equation from unpublished work by Schwanhäusser et al.). Because iBAQ is integrated into the software package MaxQuant (Cox and Mann, 2008), which is also capable of processing SILAC data, combining absolute and relative quantification is straightforward. iBAQ has been applied in various works to quantify mammalian protein absolute abundances (Schwanhäusser et al., 2011; Nagaraj et al., 2011; Geiger et al., 2012).

Although label-free quantification has been broadly applied it is extremely sensitive to variability in sample preparation (pipetting errors, incomplete digestion, inaccurate sample injection), LC-MS reproducibility, and ionization efficiency; any difference in sample handling can lead to measurement errors and affect the reliability of quantification. Therefore chromatographic peak alignment, peak matching, data normalization and statistical analysis should be applied to gain insight from the data and avoid inaccuracies in quantification (Chelius and Bondarenko, 2002; Ono et al., 2006; Choi et al., 2008). Also, care should be taken if any fractionation will be applied to the samples. For example it has been observed that in SDS-PAGE fractionation loss of peptides due to low recovery from the gel may further affect the quality of label-free quantification (Havliš and Shevchenko, 2004). Low recovery from the gel has been explained by insufficient digestion process (Havliš and Shevchenko, 2004; Getie-Kebtie et al., 2011). In another study SDS-PAGE separation was demonstrated to moderately decrease the reproducibility of quantification while improving the proteome coverage (Gautier et al., 2012).

2.3.3 *Labeled versus label-free quantification*

Several studies have been carried out in order to compare different quantification approaches. Because label-free quantification is an easy and a cheap alternative to approaches using labeling techniques, this overview will concentrate on studies comparing those two methods.

Spectral counting quantification of *Saccharomyces cerevisiae* membrane fractions was compared with peak intensity based quantification by metabolic labeling with ^{15}N (Zybailov et al., 2005). Strong correlation was found between two label-free methods when high abundant peptides were compared (Pearson's correlation of 0.64 (R^2 = 0.41) for all 645 quantified proteins), however, spectral counting was reported to be more reproducible and have a wider dynamic range than peak intensity based methods (Zybailov et al., 2005). In another study, label-free methods performed better than chemical labeling ICAT applied on standard proteins and when compared to each other in *Francisella novicida* cell lysate, peak intensity based quantification outperformed spectral counting (Ryu et al., 2008).

A study of *Methanococcus maripaludis* proteome quantification with ^{15}N metabolic labeling or with spectral counting (R_P = 0.57 (R^2 = 0.32)) revealed that although spectral counting quantification had a wider dynamic range, it was less sensitive to detecting small changes ($<$ 2 times) (Hendrickson et al., 2006).

Spectral counting in combination with MS/MS peak intensity measurements provided a higher dynamic range (up to changes 1:60) than regular spectral counting and SILAC methods for screening phosphotyrosine binding proteins in HeLa cells (Asara et al., 2008).

Peak intensity methods based on the three most intense peptides and DIA mode outperformed chemical labeling iTRAQ (R_P = 0.69 (R^2 = 0.48)) in terms of analysis time and number of proteins quantified in a *Methylocella silvestris* cell lysate (Patel et al., 2009).

2D gel electrophoresis and APEX were compared for global quantification of *Shigella dysenteriae* proteome and were found to have reasonably good correlation (R^2 = 0.67) for 255 protein quantities detected by both methods (Kuntumalla et al., 2009).

A comparison of SILAC and spectral counting quantification of human embryonic stem cells showed low correlation between the methods and that spectral counting provided less precise quantification of proteins with low number of spectral counts (Collier et al., 2010).

Spectral counting, metabolic labeling with ^{15}N and isobaric chemical labeling iTRAQ and TMT were systematically compared using *Pseudomonas putida* cell lysate and found that spectral counting covers more proteins and has larger dynamic range but is less reproducible than labeling approaches; chemical labeling is more precise and reproducible than metabolic labeling (Li et al., 2011).

While these examples are not exhaustive, we can conclude that comparing label-free methods with isotope labeling methods is a popular topic. So far it has been proven that advantages of label-free methods, additionally obvious ease of use and low cost, are also larger dynamic rage (Zybailov et al., 2005; Hendrickson et al., 2006; Asara et al., 2008) and better proteome coverage (Collier et al., 2010; Li et al., 2011). Better proteome coverage by label-free quantification has been explained by less MS analysis time spent in the label-free approach on redundant peptides that differ only in the number of isotopes (Li et al., 2011). Disadvantages of label-free methods are sensitivity to any deviation in paralleled sample preparation and low sensitivity to small protein changes between different samples (Hendrickson et al., 2006; Asara et al., 2008).

2.3.4 *From relative to absolute quantification*

The ultimate purpose of quantitative proteomics is the measurement of absolute protein abundances. Absolute concentrations of proteins are expressed as the amount of each protein per unit of biomass — for example, molecules per cell — and they can be used independently to characterize the amount of different proteins in the sample. Absolute quantification provides a more precise description of molecular events in the biological processes than relative quantification (Vogel and Marcotte, 2012). Knowledge of absolute levels of proteins in cells is important for kinetic modeling of biological processes (Tolonen et al., 2011), calculation of protein half-lives (Schwanhäusser et al., 2011), determination of the stoichiometry of protein complexes (Kuntumalla et al., 2009; Maier et al., 2011), or comparison of concentration differences between proteins within or across samples or species (Ludwig et al., 2012; Nagaraj et al., 2011; Geiger et al., 2012). As explained above, this overview considers all methods that report protein amounts, rather than ratios between different states, as an absolute quantification.

As absolute quantification of full proteome with isotope labeled standards remains technically challenging (Picotti et al., 2009, 2010) label-free quantification is used instead. Absolute concentrations measured using label-free quantification are estimated either by splitting the total amount of protein in the sample among all proteins identified (Ishihama et al., 2005; Lu et al., 2007; Nagaraj et al., 2011) or absolute abundances for the whole proteome are estimated based on linear regression of standard proteins (Ishihama et al., 2008; Schwanhäusser et al., 2011; Maier et al., 2011; Schmidt et al., 2011).

The total amount of protein to split over the quantified proteins has been calculated by florescence absorbance (Nagaraj et al., 2011), or applied based on textbook knowledge (Lu et al., 2007), which is not very precise. Even worse, the measurement of proteomes by mass spectrometry methods are limited in the number of identified proteins, and therefore, this type of abundance estimation is a very rough indication of actual concentration.

If linear regression is used, accurate absolute protein abundances are determined only for a small number of proteins spanning the whole protein abundance range. The Sigma-Aldrich® universal proteomics standard (UPS2), which consists of 48 accurately quantified human proteins formulated into a dynamic range of concentrations spanning six orders of magnitude, was used to form a calibration curve in order to quantify absolute protein concentrations in a mammalian cell line (Schwanhäusser et al., 2011). In another approach, three MS based quantification methods were combined: SRM analysis of isotope labeled reference peptides to quantify small number of anchor proteins, median intensity of the top three most intense peptides, and spectral counting (Malmström et al., 2009; Maier et al., 2011; Schmidt et al., 2011; Beck et al., 2011) to estimate protein abundances to a complete proteome of *Leptospira interrogans* (Malmström et al., 2009; Schmidt et al., 2011), *Mycoplasma pneumoniae* (Maier et al., 2011) and human cell line (Beck et al., 2011).

To gain further insight from label-free quantification and improved sensitivity of relative quantification by metabolic labeling, these methods can also be used in parallel. Absolute protein concentrations in a proteome-wide analysis of a biofuel-producing microbe *Clostridium phytofermentans* were estimated using APEX; relative changes between treatments were quantified by chemical dimethylation with stable isotope-enriched reagent (Tolonen et al., 2011).

Wide dynamic range of protein abundances in proteomes have been detected with different MS approaches: 2-2,000 copies per cell in *M. pneumoniae* (Maier et al., 2011); 1-40,000 copies per cell in *L. interrogans* (Malmström et al., 2009); 100-10^5 copies per cell in *E. coli* (Ishihama et al., 2008); 10-10^6 copies per cell in yeast (Picotti et al., 2009), 10-10^7 copies in mouse fibroblast (Schwanhäusser et al., 2011); and 500-2×10^7 in a human cell line (Beck et al., 2011). This remarkable difference in protein abundance between organisms indicates that the dynamic range of the proteome correlates with both the size of the cells and their genome. Also, the proteome coverage with MS based analysis is more or less correlated with the size of the organism. The most comprehensive proteome coverage achieved so far is 74% for *M. pneumoniae* (Maier et al., 2011); 51% for *Leptospira interrogans* (Malmström et al., 2009); 60% for *E. coli* (Iwasaki et al., 2010); 63% for yeast (Nagaraj et al., 2012); and ~50% for a mammalian cell line (Geiger et al., 2012). There are also remarkable differences between the abundances of protein functional groups between mammalian cells and microorganisms: for example in mammalian cells half of

the protein mass is devoted to regulatory mechanisms, while in microorganisms only 25% (Beck et al., 2011).

2.4 RELATIONSHIPS OF PROTEOMICS WITH OTHER –OMICS METHODS

The central dogma of molecular biology was formulated by Francis Crick in 1958 and it involves the principle of unidirectional flow of information from DNA to messenger RNA (mRNA) to the resulting proteins (Crick, 1970). This is realized in two processes: transcription and translation. Transcription is a process in which mRNA molecules are synthesized from DNA templates and translation is a process in which proteins are synthesized from mRNA templates. In order to understand microbial metabolism and its responses to environmental changes, all gene products should be measured: mRNA, proteins and metabolites. mRNA expression alone does not provide information of the amount of protein produced, its location, activity, or functional relationship with metabolites (Zhang et al., 2010). Gene expression regulation can take place at transcriptional, post-transcriptional, translational or/and post-translational level. It remains challenging to combine different –omics methods to understand the gene expression regulation levels (Maier et al., 2011; Schwanhäusser et al., 2011). While previous studies have concentrated on the comparison of protein-mRNA pairs (see below), new methods using metabolic labeling are also able to measure protein half-lives. The analysis of the dynamics of mRNA-protein pairs over 4 days in batch culture revealed that in *M. pneumoniae*, these dynamics are dominated by translational regulation (Maier et al., 2011). Protein and mRNA abundances, together with corresponding half-lives, revealed that in mouse fibroblasts, gene control is also controlled at the translation level (Schwanhäusser et al., 2011).

2.4.1 *Transcription units in bacteria*

Most of the bacterial genes are organized into polycistronic operons (Lodish et al., 2000) and, under various conditions, operons could be divided into smaller transcriptional units, resulting in many alternative transcripts (Güell et al., 2009). It was thought until recently that this organization into operons or transcription units leads to an equal level of expression of all the genes in the units (Laing et al., 2006). However, a recent genome-wide transcriptomics study revealed that in *M. pneumoniae* consecutive genes within the operon did not have the same expression level, leading to operon polarity (Güell et al., 2009). Almost half of the consecutive genes showed staircase-like decay, meaning that the consecutive genes have lower and steady expression levels following 5′ to 3′ directionality (Güell et al., 2009). The staircase behavior was also reported for *Streptomyces coelicolor* but in the same study not found for *E. coli* operons (Laing et al., 2006).

Operon polarity could be explained by the combinatorial effect of internal promoters and terminators that would result in the production of different transcripts from the same operon, therefore leading to unequal expression levels of gene products (Güell et al., 2011). At the proteome level, decay of consecutive genes has been described in antibiotics treated *E. coli* ribosomal protein operons, similar to staircase behavior (Siibak et al., 2011), however, a gradual decrease in protein production was clearly seen only for a few

non-ribosomal operons. Maier et al. (2011) stated, after comparing mRNA-protein profiles in operons, that the previously observed operon polarity of consecutive transcripts in *M. pneumoniae* (Güell et al., 2009) tends to be compensated on the protein level and Schmidt et al. (2011) discovered staircase-like behavior at the proteome level for only 5% of *L. interrogans* operons. Therefore, there is no clear evidence that staircase behavior takes place also at the proteome level.

2.4.2 *mRNA and protein level correlation*

Transcriptomics, also called global analysis of gene expression or genome-wide expression profiling, has been one of the tools to measure all mRNA molecules, or "transcripts", produced in one cell or a population of cells (Zhang et al., 2010). Before proteomics methods became widely used, mRNA concentrations were used to express the corresponding proteins concentrations, assuming that transcript abundances were the main determinants of protein abundances (Vogel and Marcotte, 2012). However, technological advances in mass spectrometry have allowed to carry out large-scale studies of proteomes and correlate these with their transcriptomes. In general, in both bacteria and eukaryotes, protein levels correlate with their corresponding mRNA levels, but not very strongly. The squared Pearson's correlation coefficient in the range of ~0.40-0.60 have been observed (Lu et al., 2007; Maier et al., 2009; Schwanhäusser et al., 2011), which implies that only about 40-60% of the variation in protein concentration can be explained by knowing mRNA abundances (de Sousa Abreu et al., 2009; Vogel and Marcotte, 2012). This low correlation can be explained as processes involved in gene expression involve besides mRNA and protein synthesis also degradation rates of mRNAs and proteins (Vogel and Marcotte, 2012). Therefore, there is no one-to-one relationship between protein amounts and their corresponding mRNAs.

Low correlation between relative mRNA and protein levels has been reported for several organisms such as *Saccharomyces cerevisiae* (Spearman rank correlation $S_r = 0.21$) (Griffin et al., 2002), *Halobacterium* (Baliga et al., 2002), human cancer cells (Chen et al., 2002) and *Lactococcus lactis* (Dressaire et al., 2009). Weakly positive correlation ($S_r = 0.45$) for absolute abundances was demonstrated for yeast (Washburn et al., 2003). Moderate correlation between protein and mRNA absolute abundances has been reported for *Plasmodium falciparum* (Le Roch et al., 2004) (S_r up-to 0.59). A delay between mRNA and protein accumulation, based on comparative analysis of relative expression data was observed in the studies of *Plasmodium falciparum* (Le Roch et al., 2004).

As proteomics and transcriptomics technologies have improved, the correlation between mRNA and proteins has also improved, probably by chance, however, the Spearman rank correlation has been mostly replaced with Pearson's squared correlation coefficient, which we will refer from here on. The following Pearson's squared correlation coefficients between absolute abundances of proteins and mRNA have been reported: 0.73 for *Saccharomyces cerevisiae* (Lu et al., 2007), 0.47 for *Escherichia coli* (Lu et al., 2007), 0.40 for *Streptomyces coelicolor* (Jayapal et al., 2008), 0.27 for *Mycoplasma pneumoniae* (Maier et al., 2011) and 0.41 for mouse fibroblasts (Schwanhäusser et al., 2011).

The poorer correlation between mRNA and protein expression levels in early studies could be partly explained by technical limitations. 2D gel quantification was affected by signal saturation and multiple proteins per spot (Lu et al., 2007). Also the low number of proteins (Gygi et al., 1999b; Griffin et al., 2002), or protein and mRNA data from different experiments or conditions were correlated (de Groot et al., 2007), which makes the overall correlation unreliable. It has been proposed that the greatest source for the variation of the mRNA-protein correlation is the variation of mRNA or protein abundance measurements (Nie et al., 2006). Variation in protein abundance contributed 34-44% and mRNA 9-22% to the variation of mRNA-protein correlation (Nie et al., 2006). However, variations in measurements are strongly dependent on the experimental setup and equipment used. Obtaining quantitatively reliable data on the proteome (and transcriptome) is one of the most important challenges in systems biology in order to understand organism-specific regulation of translation and protein degradation (Maier et al., 2011; Schwanhäusser et al., 2011; Boisvert et al., 2012; Martin et al., 2012).

2.4.3 *Integrating metabolomes and proteomes*

Metabolites are small molecules within cells, and their concentration levels vary as a consequence of genetic or physiological changes. Metabolomics is a method to measure the diversity and abundances of the metabolites in the cell (Raamsdonk et al., 2001). To date, there is no known direct relationship between cellular metabolite concentrations and gene expression as occurs between mRNA and protein levels (Zhang et al., 2010).

Measuring the concentrations of substrates and extracellular by-products together with biomass amounts and composition allows one to calculate quantitative input-output flux values, as is accomplished using metabolic flux analysis (MFA) (Valgepea et al., 2011). Metabolic flux analysis, together with quantitative proteome measurements, allows one to understand the mechanisms of how enzymes are regulated. Metabolic regulation through the transcription and translation processes takes time and is inadequate for cells to cope with a rapidly changing environment. For efficient metabolic regulation, enzyme activities are regulated (Kim and Gadd, 2008). A targeted study of 228 protein abundances together with metabolic flux analysis (MFA) in the central carbon and amino-acid metabolic network of *Saccharomyces cerevisiae* revealed that newly required fluxes are regulated by protein abundance (Costenoble et al., 2011), while changes in already existing fluxes are probably regulated by enzyme activities.

Maximal enzymatic activity can be measured using *in vitro* enzyme assays (Canelas et al., 2010; Tolonen et al., 2011), however, those values do not reflect the enzyme activity in the cell under specific experimental conditions.

2.5 MICROBIAL CULTIVATION METHODS

The simplest microbial cultivation method is batch cultivation, and is usually carried out in a shake-flask, where nutrients are not added after inoculation. Batch culture is a heterogeneous system, where the environment continuously changes as growth, product formation and substrate utilization take place all at the same time (Hoskisson and Hobbs, 2005).

However, in order to collect reproducible and meaningful information with regards to cell physiology, the culture must be grown under controlled conditions (Hoskisson and Hobbs, 2005). The simultaneous development of the continuous culture system chemostat by Monod (Monod, 1950) and Novick & Szilard (Novick and Szilard, 1950) allowed microbial physiologists to study bacterial growth under constant physiochemical environments.

Chemostat begins as batch culture until it reaches the exponential growth phase. After that, fresh medium is added and the same amount of culture is removed at a steady suitable rate, which allows the cells to grow at a specified steady rate. Environmental parameters, pH, temperature, nutrients, and metabolic products can all be varied and controlled. In chemostat cultivation cells are in a "steady state" (Hoskisson and Hobbs, 2005), meaning that all the cells are growing with the same growth rate (μ, h^{-1}) which is equal to the dilution rate (D, h^{-1}) (Novick and Szilard, 1950; Monod, 1950). Steady state is achieved after the culture is stabilized during the flow-through of 4-5 culture volumes.

Accelerostat (A-stat) is a modification of chemostat, where the specific growth rate is changed with smooth acceleration, so that the specific growth rate remains equal to the dilution rate (Paalme et al., 1995). This method has advantages over chemostat, because it enables one to collect data at different growth rates in one experiment, thus saving time. With A-stat culture it is possible to precisely detect metabolically relevant switch points, for example start of overflow metabolism, which could have been unnoticed using chemostat culture (Adamberg et al., 2009; Valgepea et al., 2010).

Most proteomics studies with cultivable cells are performed using batch cultivation (Malmström et al., 2009; Maier et al., 2011; Tolonen et al., 2011; Nagaraj et al., 2012), however, there are also studies where chemostat cultivation has been applied for global proteome characterization (Rathsam et al., 2005; Kolkman et al., 2006; de Groot et al., 2007; Ishii et al., 2007; Dressaire et al., 2009).

It has been claimed that batch cultivation complicates mRNA-protein correlation while the reproducibility of growth rate under controlled steady environmental conditions are likely to result in more accurate quantification (Kolkman et al., 2006). However, the correlation between proteins and mRNAs in chemostat cultivated *Saccharomyces cerevisiae* remained moderate (S_r = 0.55) for 285 pairs (Kolkman et al., 2006). The only example of proteome quantification in an A-stat cultivation is a growth rate dependent study of *Lactococcus lactis* (Lahtvee et al., 2011); R^2 = 0.48 was observed between 600 relatively quantified protein and gene pairs, when comparing cells growing with fast or slow growth rate (Lahtvee et al., 2011). Systems biology is used to study global transcriptomics, proteomics, metabolomics, along with several other –omics methods and requires reproducible, reliable and biologically homogeneous datasets to quantitatively characterize the physiological states of cells or organisms. The use of continuous culture techniques such as chemostat and A-stat are important tools in the acquisition of such data. Steady state continuous culture techniques are preferred over batch culture, where heterogeneous growth and stress, caused by accumulating metabolic products, can often mask subtle physiological differences and trends (Hoskisson and Hobbs, 2005). Controlled culture conditions are an important issue for the reproducibility of experiments and therefore continuous cultivation techniques should be applied more often in systems biology studies (Schaechter, 2006).

3

AIMS OF THIS DISSERTATION

This dissertation had three main aims:

I Development of methods for the quantitative analysis of a growth rate dependent *E. coli* proteome,

II Comparison of different quantification methods to characterize the *E. coli* proteome,

III To further our understanding of *E. coli* metabolism using proteome, transriptome, metabolome and cultivation data.

4

MATERIALS AND METHODS

More detailed descriptions of the materials and methods applied are available in the publications. The following sections are provided to make this material more accessible.

4.1 BACTERIA CULTIVATION

E. coli K-12 MG1655 (λ- F- *rph-1Fnr+*; Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), DSM No.18039) was cultivated on glucose minimal medium in A-stat culture, as described by Valgepea et al. (2010). Samples for proteome analysis were collected from specific growth rates 0.10 h^{-1} (chemostat point prior to the start of acceleration in A-stat); 0.21; 0.30; 0.40; 0.49 h^{-1}. Samples were collected from the fermenter, washed with PBS (0.137 M NaCl, 2.7 mM KCl, 10.0 mM Na_2HPO_4, 1.4 mM KH_2PO_4), flash frozen in liquid nitrogen and stored at -80℃ prior to protein extraction.

4.2 SAMPLE PREPARATION

4.2.1 *SDS-PAGE*

Proteins were extracted and total amount of protein was quantified. For metabolic labeling with ^{15}N labeled standard (Publication II), culture was mixed in a 1:1 ratio with samples from continuous cultivation. Samples were separated on a SDS-PAGE, lanes were cut to 10-40 slices and in-gel digested. Details are provided in Publication I and Publication II.

4.2.2 *Shotgun*

Proteins were extracted, and the total amount of protein was quantified. The Universal Proteomics Standard (UPS2, Sigma-Aldrich) was added to samples in 1:3 ratio, followed with overnight in-solution digestion, as described in Publication III.

4.3 MASS-SPECTROMETRY

4.3.1 *"In-source CID" in LCT Premier*

Orthogonal-acceleration time-of-flight instrument oa-TOF-MS LCT Premier (Waters, UK) was coupled with UHPLC (Waters, UK). A peptide mixture (~0.25 µg) was loaded on a column (Waters ACQUITY UPLC® BEH300 C18 1.7 µm, 2.1 mm x 100 mm) using a 20 minute gradient from 5% to 40% solvent B (solvent A: MilliQ H_2O/0.1% formic acid, solvent B: 100% acetonitrile/0.1% formic acid) and a flow rate of 0.15 $mL \cdot min^{-1}$. LCT Premier mass spectrometer operated in positive mode, mass range 300-2,000 Da, resolution up to 10,000 FWHM, at 2 kV capillary and 30 V sample cone voltage, 300°C desolvation and 120°C source temperature. LCT Premier was run simultaneously in two modes, switching every second between low and high energy (15 and 55 V, respectively) in the aperture between ion guides, resulting in data independent acquisition where all the precursor ions entering the source were fragmented. Data was collected with MassLynx 4.1 software (Waters, UK).

4.3.2 *QStar Elite*

A Q-TOF mass spectrometer QStar Elite (Applied Biosystems/MDS SCIEX, Germany) was connected to a P680 chromatographic pump (Dionex, Sunnyvale, CA) equipped with a degasser (Dionex Corporation, Sunnyvale, USA). The pump was fitted with an in-house built splitter, where the flow rate of 100 $\mu L \cdot min^{-1}$ was split before the column to 10 $\mu L \cdot min^{-1}$. Peptides were separated on a PepSwift monolithic PS-DVB column (500 µm I.D.×50 mm, Dionex). A 100 minute linear gradient was used from 0% to 60% solvent B (solvent A: MilliQ H_2O/0.1% formic acid, solvent B: 80% acetonitrile/0.1% formic acid) at flow rate of 10 $\mu L \cdot min^{-1}$. Mass spectra were acquired in positive ion mode setting the spray voltage at 1.80 kV, the capillary temperature at 250°C and the voltage of tube lens at 140 V. Mass spectra (300-2,000 Da) and tandem mass spectra were recorded in positive-ion mode with a resolution of 10,000-12,000 FWHM. Data acquisition was performed with an ion spray voltage of 5.5 kV, declustering potential of 60 V and focusing potential of 320 V. Data dependent acquisition was used to obtain tandem mass spectrometry (MS/MS) spectra for the five most intensive peaks following each MS survey scan.

4.3.3 *LTQ Orbitrap*

An LTQ Orbitrap mass-spectrometer (Thermo Electron, Bremen, Germany) was connected to an Agilent 1200 series nanoflow system (Agilent Technologies, Santa Clara, CA). Peptides were loaded on a self-packed fused silica emitter (150 mm x 0.075 mm; Proxeon, Denmark, packed in-house with ReproSil-Pur C18-AQ 3 µm particles (Dr. Maisch, Germany)) using a flow rate of 0.7 $\mu L \cdot min^{-1}$ and 90 or 120 minute gradient for SDS-PAGE separated samples (Publication II) and 240 minute gradient for shot-gun samples (Publication I). For all methods, the gradient was from 3% to 40% solvent B (solvent A: MilliQ H_2O/0.1% formic acid, solvent B: 80% acetonitrile/0.1% formic acid) with a flow-rate of

0.2 $\mu L \cdot min^{-1}$. Peptides were sprayed directly into an LTQ Orbitrap mass-spectrometer operated at 180°C capillary temperature and 2.4 kV spray voltage. One full mass spectra was acquired in FT-ICR profile mode, with a mass range 300-1,900 Da at a resolving power of 60,000 FWHM, following by data-dependent fragmentations of the five most intense multiply charged ions acquired in centroid mode in the linear ion trap.

4.4 DATA ANALYSIS

4.4.1 *"In-source CID" data analysis*

Fragmented peptide spectra were aligned manually with information on their precursor masses and charges and searched by Mascot search engine using NCBI *E. coli* database. Details are provided in Publication I.

4.4.2 *Metabolic labeling with ^{15}N*

Peak lists for database searches were produced with Raw2MSM and searched by Mascot search engine against *E. coli* K-12 MG1655 protein sequence database (`http://ecogene.org`). Quantification of $^{15}N/^{14}N$ ratios was performed using the MSQuant program. Two biological replicates were compared at specific growth rates 0.20; 0.26; 0.30; 0.40; 0.49 h^{-1} with sample at μ = 0.10 h^{-1} (chemostat point after stabilizing the culture in A-stat). Details are provided in Publication II.

4.4.3 *Spectral counting based absolute quantification*

Exponentially Modified Protein Abundance Index (emPAI) values were obtained directly from the Mascot database search results. Absolute protein expression indexes (APEX) were calculated from Mascot search results processed with the APEX Quantitative Proteomics Tool.

The total concentration of protein copies per cell (2×10^6) at a specific growth rate of 0.11 h^{-1} was calculated based on biomass concentration, cell size and total protein concentration. Total protein copies per cell was used as a normalization factor to determine individual protein copies per cell for all identified proteins from the emPAI and APEX indexes. Details are provided in Publication III.

4.4.4 *Intensity-based absolute quantification (iBAQ)*

Raw data files were analysed with the MaxQuant software and the protein concentrations in fmol were calculated based on linear regression of UPS2 standard proteins. Protein copies per cell were calculated by multiplying the molar concentration with Avogadro constant and dividing with the number of cells in the respective experiment obtained by plate counting (8-9 $\times$ 10^9 cells$\cdot mL^{-1}$) (Valgepea et al., 2010). Details are described in Publication III.

4.4.5 *Label-free quantitative data analysis*

Absolute quantitative data for different growth rates was calculated using protein abundance at growth rate 0.1 h^{-1} and relative ratios (0.2/0.1, 0.3/0.1, 0.4/0.1, 0.49/0.1 h^{-1}) collected from ^{15}N metabolic labeling experiments.

All identified and quantified proteins were grouped into Clusters of Orthologous Groups (COGs) and divided into transcription units and functional complexes according to the EcoCyc database (Keseler et al., 2011) using the in-house built script.

For correlation analysis, Analyse-it for Microsoft Excel (version 2.20) was used. Pearson's correlation test was applied to biological replicates as well to samples analysed with different quantitative proteomics methods and proteomics and mRNA correlation.

Variability was characterized by the coefficient of variation (CV, %), which is defined as the ratio of the standard deviation to the arithmetic mean.

Staircase-like expression was analysed for transcription units with two or more components. Protein expression levels of transcription units were sub-divided into "no staircase" and staircase-like behaviour types "up", "down" and "others". A transcription unit was classified as "no staircase" if at least half of its consecutive genes were not differentially expressed. Two consecutive genes in the transcription unit were considered differentially expressed if their protein abundance measurements for two biological replicates did not overlap. Staircase-like behaviour expression of transcription units was classified as "up" or "down" if at least half of its consecutive genes were differentially expressed at higher or lower levels, respectively, in the mRNA emerging direction during transcription ($5' \rightarrow 3'$). The remaining transcription units were classified as "others".

The cost of protein synthesis was calculated by multiplying the respective proteins' abundance in the cell with its peptide bond count (number of amino acids in the protein minus one) and 4.306 ATP (Stouthamer, 1973) which stands for the cost in ATP for one amino acid polymerization reaction in the ribosome.

Apparent enzyme activities (k_{cat}, s^{-1}) per protein chain or subunit (without taking into account the number of proteins and catalytic sites necessary for catalytic activity) were calculated as follows:

$$k_{cat} = \frac{\text{specific flux} \times N_A}{\text{iBAQ}},$$

where specific flux ($mmol \cdot g_{DCW}^{-1} \cdot h^{-1}$; g_{DCW} — grams of dry cellular weight) was obtained in the same experiments for respective reaction and published previously (Valgepea et al., 2010), N_A is Avogadro constant, iBAQ is a protein abundance (protein copies in g-DCW). In vitro enzyme assays measure maximal enzymatic activity (Tolonen et al., 2011); however, those values do not reflect the enzyme activity in the cell at certain experimental conditions. In our studies we used apparent k_{cat} values of enzymes, which were defined as average throughput of molecules per protein chain catalyzing given reaction in an experiment. Apparent k_{cat} can be calculated for each enzyme at specific experimental condition as a ratio of specific flux and protein abundance designated for the respective flux. Apparent k_{cat} calculations were based on absolute amounts of 266 enzymes and 66 metabolic fluxes in the main metabolic network; covering glycolysis, the tricarboxylic

acid cycle, pentose phosphate pathway, respiratory chain, and biopolymer monomer synthesis.

For gene regulation analysis covariance coefficients were calculated: 1) between protein/mRNA and specific growth rates; 2) between flux/protein *i. e.*, k_{cat} and specific growth rates (Table 1). Uncertainty values were calculated for covariance analysis for testing the statistical hypothesis of covariance values being statistically different from zero. Calculated covariance coefficients were subjected to statistical hypothesis testing: one sided t-test was applied to test the hypothesis that absolute values of covariance are higher than zero at statistically significant level. Covariance value was considered to be zero if the p-value was below 0.05. Genes were divided into three groups: 1) genes with covariance value statistically higher than zero; 2) genes with covariance value equal to zero at statistically significant level; 3) rest of the genes that are described by very high uncertainty level of covariance — in this case no regulation analysis could be applied. The gene was considered as transcriptionally (TR) regulated if the covariance coefficient between protein/mRNA and specific growth rate was zero, *i. e.*, protein/mRNA remained constant at all growth rate values; post-transcriptionally (P-TR) regulated if there was negative covariance between protein/mRNA, *i. e.* less proteins produced per mRNA with an increase in specific growth rate and translationally (TL) regulated if there was positive covariance between protein/mRNA and specific growth rate. For 266 genes, where also the apparent k_{cat} was calculated, additional analysis, including covariance between k_{cat} and the specific growth rate, was applied. Post-translational (P-TL) regulation of some of the enzymes previously considered as transcriptionally regulated was observed — if k_{cat} values were statistically proven to depend on the specific growth rate (Table 1).

Table 1 – Rules for gene regulation analysis. COVARIANCE – Covariance analysis between the protein/mRNA ratio or k_{cat} and specific growth rate. Statistical tests were performed to differentiate, which protein/mRNA (pm) or k_{cat} values were statistically unchanged ("$= 0$") or which were statistically different from zero ("$\neq 0$"). Regulation patterns for gene products that did not pass these two statistical tests were not considered in the gene regulation analysis.

COVARIANCE with μ	Regulation
$\Delta\text{protein}/\Delta\text{mRNA} = 0$, $\Delta k_{cat} = 0$	Transcriptional
$\Delta\text{protein}/\Delta\text{mRNA} < 0$, $\Delta k_{cat} = 0$	Post-transcriptional
$\Delta\text{protein}/\Delta\text{mRNA} > 0$, $\Delta k_{cat} = 0$	Translational
$\Delta\text{protein}/\Delta\text{mRNA} < 0$, $\Delta k_{cat} \neq 0$	Post-transcriptional / Post-translational
$\Delta\text{protein}/\Delta\text{mRNA} > 0$, $\Delta k_{cat} \neq 0$	Translational / Post-translational
$\Delta\text{protein}/\Delta\text{mRNA} = 0$, $\Delta k_{cat} \neq 0$	Transcriptional / Post-translational

5

RESULTS AND DISCUSSION

The results of this dissertation are presented and discussed in the following sections.

5.1 "in-source CID" applied in proteomics (Publication I)

A method for peptide identification using a relatively low-cost ESI TOF mass-analyzer LCT Premier (Waters) was developed and tested. Simultaneous accurate precursor ion measurement and peptide fragmentation was performed, using an "in-source CID" which occurs in LCT Premier MS in the region between the first ion guide and the aperture separating the ion guides. In the first ion guide region, the ions gain significant internal energy as they accelerate. Collision of these ions with neutral molecules of the residual atmospheric gas leads to their dissociation (Ren et al., 2009). In order for "in-source CID" to take place in a single MS, the analyzer must operate in two different scanning modes. In full-scan mode all ions from the source are analyzed, allowing the recording of precursor ions; in fragmentation mode all ions are subjected to an additional fragmentation step that leads to the formation of fragment ions. A similar switching approach is used for example in different versions of DIA mode: MS^E and all-ion fragmentation (AIF) (Silva et al., 2006a,b; Geiger et al., 2010a), where the tandem mass spectrometer switches between MS and MS/MS mode without any precursor ion selection. However, in tandem instruments, DIA method uses isolation, which will discard singly charged molecules, thus producing cleaner MS/MS spectra.

In Publication I, *E. coli* whole cell lysate was separated with SDS-PAGE (Figure 3A) and analyzed with LCT Premier MS working in two scanning modes collecting almost identical total ion count (TIC) chromatograms (Figure 3B). By changing the fragmentation energies, all the precursors entering the source were fragmented and both spectra were collected (Figure 3C). Unlike in a typical DDA MS/MS experiment, where precursor ions are isolated and analyzed individually, all the precursor ions were fragmented without any selection in these experiments. Hence, fragmentation spectra are much more complicated and lead to difficulties in identification of fragments. A significant drawback of the method is low automation. The peak list combination for the database search was carried out manually because it was not possible to analyze it with currently available software packages.

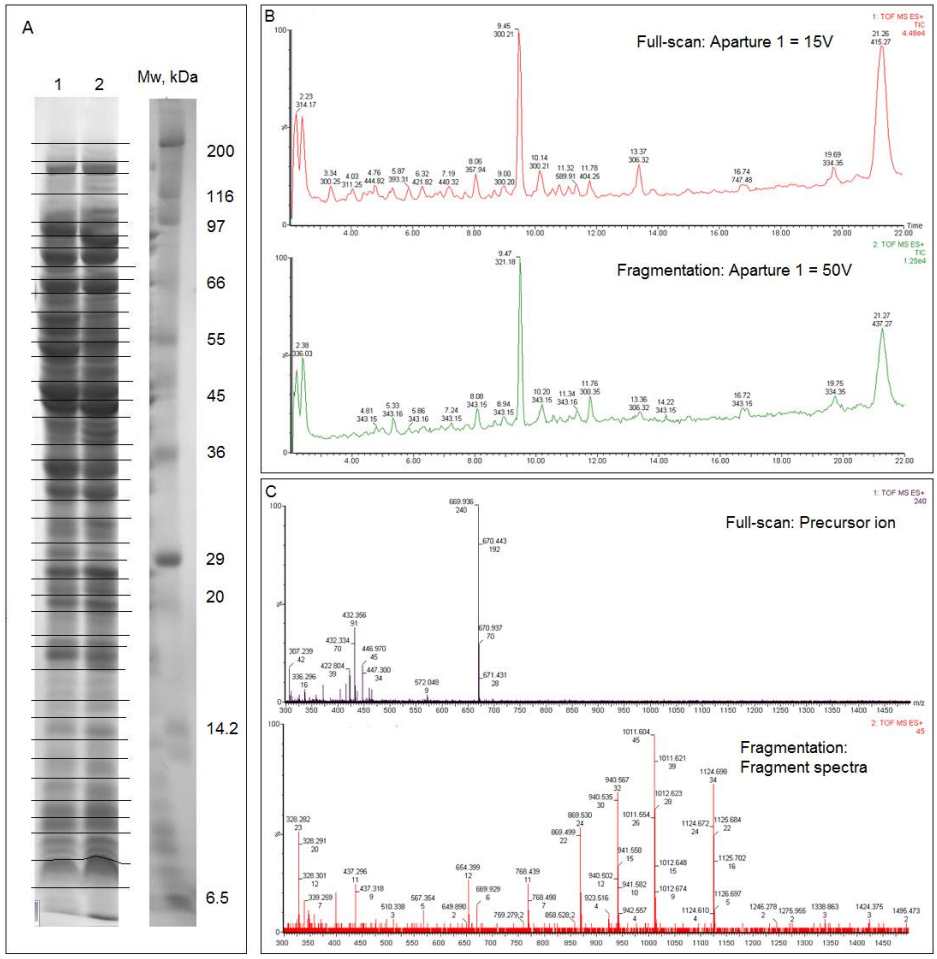


Figure 3 – (**A**) *E. coli* soluble proteins were separated by SDS-PAGE, each lane was cut into 40 fractions and fractions were digested and analyzed by LC-MS. (**B**) Two similar TIC chromatograms were collected in the same run by rapidly changing between high and low energies in the middle of ion guides. (**C**) An example of a peptide fragmented with "in-source CID". All the precursor ions which were entering the source were also fragmented, producing fragment spectra for peptide identification.

More than one hundred proteins were identified by in-source fragmentation. Two *E. coli* continuous cultivation samples were analyzed: one at specific growth rate 0.5 h^{-1} and another one at 0.2 h^{-1}. Protein identifications obtained by TOF MS were validated for six gel fractions with quadrupole-TOF MS QStar Elite (AB Sciex). All proteins, which were observed with TOF MS, were also identified with Q-TOF MS. However, three times more proteins were identified with Q-TOF MS than with TOF MS (Publication I, Fig 2a and Supplementary Table 2). This improvement with Q-TOF MS analysis can at least partly be explained by higher efficiency of automatic LC-MS/MS data analysis which was missing in

TOF MS where peak lists were collected and analyzed manually. It can only be assumed that if the data handling with TOF MS would have been automated, more peptides would be identified with TOF MS as well.

With this study we showed, as a proof-of-principle, that a single TOF MS is usable for peptide and protein identification. However, such an approach is very laborious and, because better methods exist, we did not pursue this approach further.

5.2 SET-UP OF METABOLIC LABELING IN CONTINUOUS CULTURE (PUBLICATION II)

Metabolic labeling of the growing cells is the most reliable method for quantitative proteome measurements as the label is incorporated into the cells before any sample preparation step, resulting in high reproducibility (review of Gouw et al. 2010). Usage of labeled medium is usually not cost effective in continuous cultures; therefore, stationary phase ^{15}N labeled *E. coli* K-12 MG1655 batch culture was used as a spike-in standard to study changes in *E. coli* proteome during the overflow metabolism. In this study ^{15}N labeling was used only to produce heavy labeled reference proteins, which are added to the samples from continuous cultivations after cell lysis and before protein digestion. The difference between experimental samples was calculated as the "ratio of ratios", where the ratios of samples relative to the standard were divided with each other (Figure 4).

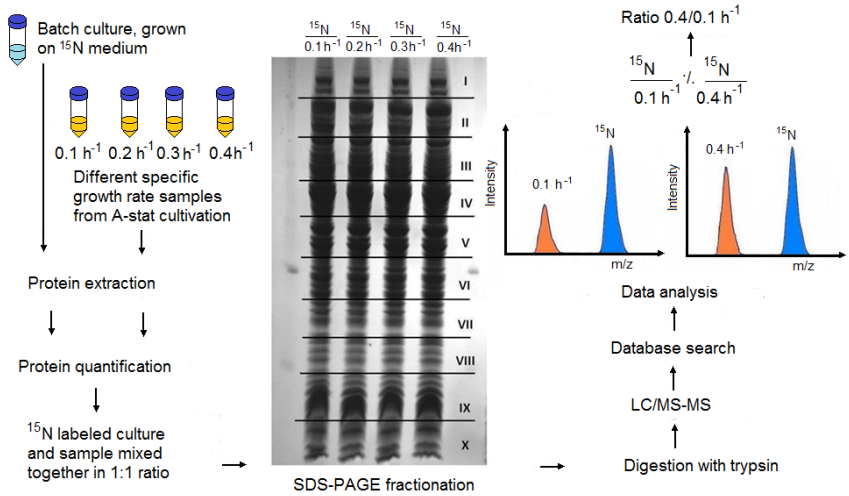


Figure 4 – The workflow of a ^{15}N-labeled experiments. The labeling is separated from the cultivation experiment, which is carried out in A-stat continuous cultivation (see Publication II). The non-labeled samples from cultivation experiment were combined with the ^{15}N-labeled standard and these combined samples were SDS-PAGE fractionated and analyzed separately by LC-MS/MS. The difference between the experimental samples was calculated as the "ratio of ratios".

As a result approximately 1,600 *E. coli* proteins, identified with at least two peptides, were quantified in two biological replicates. Protein relative abundance ratios were calculated for specific growth rates 0.2; 0.3; 0.4; 0.49 h^{-1} (A-stat samples) and compared to a sample at μ = 0.1 h^{-1} (chemostat point prior to the start of acceleration in A-stat) which produced correlation between two biological replicates in the range of R^2 = 0.62-0.84 (Figure 5). This good reproducibility for biological replicates can be explained with the controlled growth on balanced defined medium in continuous cultures (chemostat and A-stat).

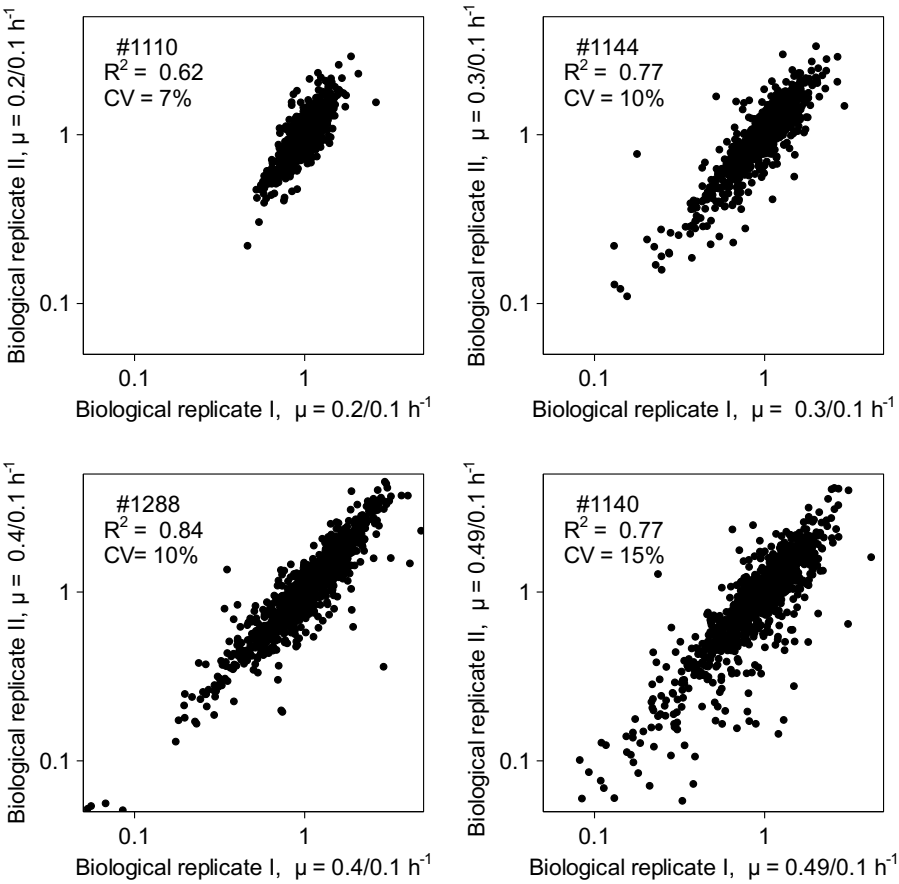


Figure 5 – Correlation between protein expression ratios of two biological replicates. Protein expression ratios were obtained by using ^{15}N metabolically labeled reference. μ – specific growth rate (h^{-1}); # – number of data points analyzed; R^2 – Pearson's squared correlation coefficient calculated for log protein expression ratios; CV – coefficient of variation in percentage. All correlations are significant at p-value < 0.0001. Axes are at log scale.

Protein expression values were calculated as the ratios of ratios (Figure 4), this leads to a decrease in the number of quantified proteins due to the missing values in one of the samples. 1,860, 1,797, 1,812, 1,738 and 1,803 $^{15}N/^{14}N$ ratios were quantified for samples at growth rates 0.1 h^{-1}, 0.2 h^{-1}, 0.2 h^{-1}, 0.4 h^{-1} or 0.49 h^{-1}, respectively. However, after calculating from those ratios new ratios to the first time point at growth rate 0.1 h^{-1}, we ended up with 1,620, 1,628, 1,595 and 1,613 values for 0.2/0.1, 0.3/0.1, 0.4/0.1 and 0.49/0.1 h^{-1} ratios, respectively. We could not mainly quantify uncharacterized proteins; however, there were unquantifiable proteins in all of the Clusters of Orthologous Groups COG functional categories (data not shown). One group of proteins which were not possible to quantify in continuous culture samples were flagella proteins, mainly due to their very low abundance in the stationary phase ^{15}N standard culture. This clearly demonstrates the limitations of relative quantification and also that care should be taken in preparing standard cells. Preferably reference cells should be collected at different growth phases and combined to a standard that has the best representative of the proteome under study, this "super-standard" approach has been applied on quantification of human tumour tissue (Geiger et al., 2010b).

5.3 ABSOLUTE QUANTITATIVE PROTEOMICS

5.3.1 *Quantification of SDS-PAGE separated proteins (Publication I and Publication III)*

It has been demonstrated that the average mass spectrometry signal response for the three most intense tryptic peptides from a protein correlates with the concentration of a current protein (Silva et al., 2006b). We applied this method in Publication I for SDS-PAGE separated proteins. First, we tested the hypothesis that the average signal response (peak area or height) of three most intense peptides should be quantitative (Silva et al., 2006b) on a standard protein — bovine serum albumin (BSA). Four different concentrations of BSA were loaded on the gel and in-gel digested (Figure 6).

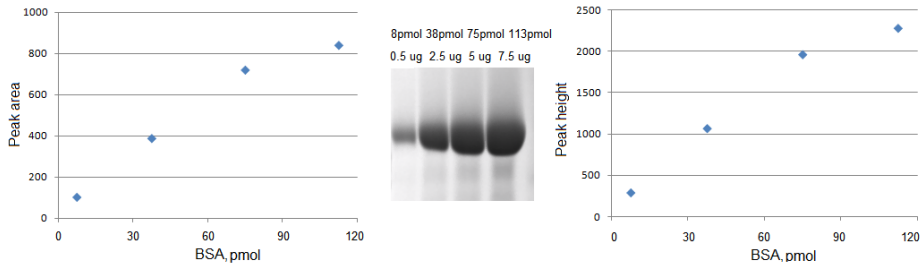


Figure 6 – Standard curves of BSA based on the MS peak areas (left) or peak heights (right) for the three most abundant peptides. The correlation is linear for peak areas or heights at BSA amount lower than 100 pmol.

The calibration curve was linear for BSA amounts lower than 100 pmol (Figure 6), the mass analyzer became saturated at higher concentrations and the linearity of response was lost. It was concluded that relative quantification can be applied between two samples extracted from the gel, based on peak area and/or height; as long as the amount of the sample remains smaller than 100 pmol or 5 µg. Peak area was used to quantify 117 gel-extracted proteins at growth rate 0.5 h^{-1} compared to the growth rate 0.2 h^{-1} (Publication I).

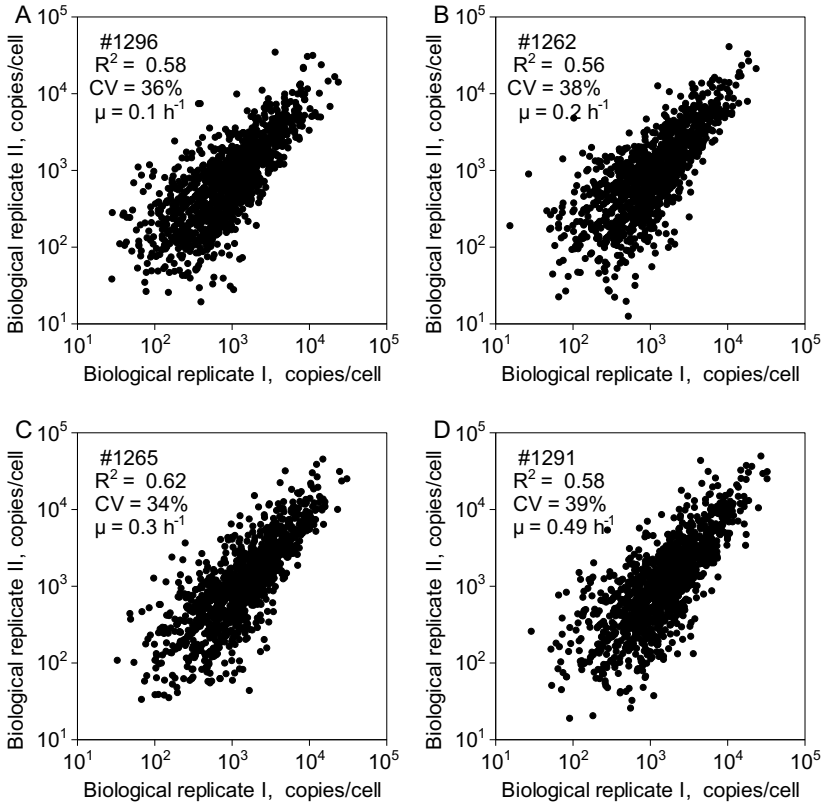


Figure 7 – Correlation between APEX protein abundances for SDS-PAGE separated and gel extracted samples of two biological replicates. µ – specific growth rate (h^{-1}); # – number of data points analyzed; R^2 – Pearson's squared correlation coefficient calculated for log protein expression ratios; CV – coefficient of variation in percentage. All correlations are significant at p-value < 0.0001. Axes are at log scale. Correlation at growth rate 0.4 h^{-1} was 0.58 for 1,263 proteins, CV 35% (data not shown).

Absolute Protein Expression (APEX) measurements were carried out on a data set of 1,808 *E. coli* proteins ($^{15}N/^{14}N$ metabolically labeled experiments, previously published in

Publication II, also including proteins based on 1 peptide identification), combined from two biological replicates separated by SDS-PAGE fractionation and extracted from the gel. The Pearson's squared correlation between two biological replicates was 0.56-0.62 (Figure 7), which is lower than for relative quantification (Figure 5). Poorer reproducibility was expected because label-free quantification is generally considered to be less accurate than metabolic labeling (Hendrickson et al., 2006; Asara et al., 2008; Collier et al., 2010; Li et al., 2011).

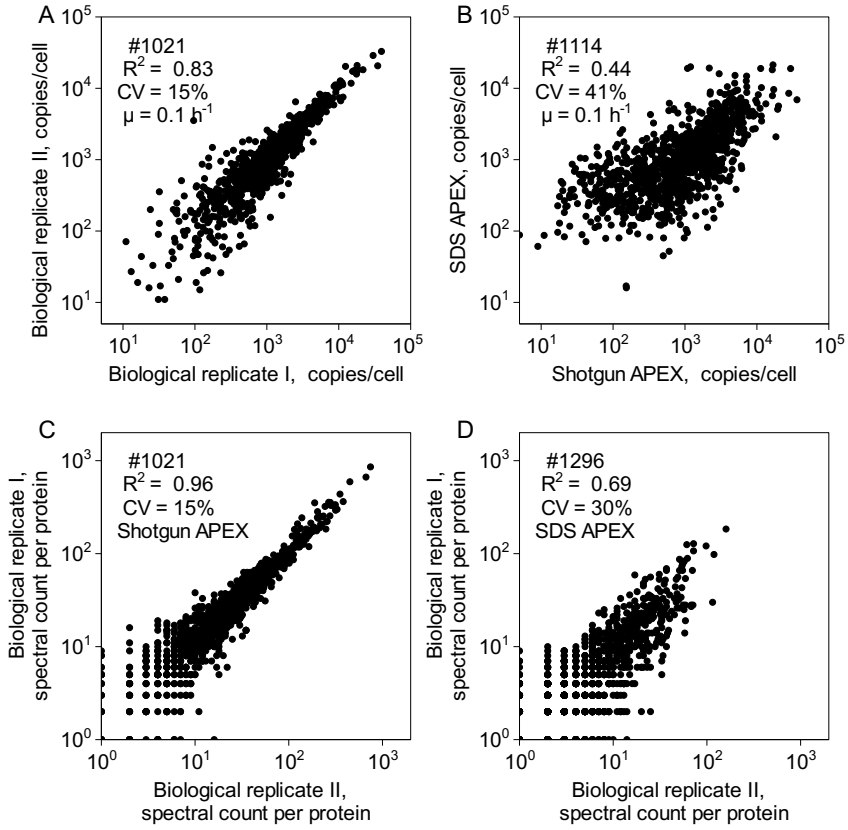


Figure 8 – The effect of sample preparation on APEX calculation. μ – specific growth rate (h^{-1}); R^2 – Pearson's squared correlation coefficient; # – number of data points analyzed. All correlations are significant at a p-value < 0.0001. **A**) Correlation between biological replicates from shotgun experiments. Axes are on a log scale. **B**) Correlation between average APEX values for samples fractionated by SDS-PAGE and no pre-fractionation (shotgun). Axes are at log scale. **C**) Correlation between spectral counts for biological replicates analyzed as a shotgun experiment. Axes are a log scale. **D**) Correlation between spectral counts for biological replicates separated by SDS-PAGE. Axes are a log scale.

In order to evaluate the effect of sample preparation on label-free absolute quantification, a sample from the specific growth rate of 0.1 h^{-1} was analyzed again by a shotgun experiment, using only a four hour nano-LC gradient for sample separation. Almost 600 proteins were detected less compared to SDS-PAGE separated sample (based on two biological replicates, data not shown). However, the correlation for 1,021 common proteins in shotgun biological replicates was found to enhance the Pearson's squared correlation to 0.83 and dynamic range improved from three to four orders of magnitude (Figure 7A compared to Figure 8A). Correlation between APEX values for SDS-PAGE fractionated and shotgun samples was $R^2 = 0.44$ (Figure 8B). Low correlation between APEX values of same samples prepared by different protocols demonstrates that label-free quantification is highly influenced by sample preparation.

Because APEX method rely on spectral counts of peptides, we also compared the spectral counts between different sample preparation methods (Figure 8C, D); less spectral counts per protein were detected for samples processed with in-gel digestion than in shotgun experiments, which indicates poor recovery of peptides from the SDS gel. It has been observed before that the digestion of proteins embedded into a polyacrylamide matrix affects the recovery of peptides (Havliš and Shevchenko, 2004), and therefore, label-free quantification may be altered if gel based methods are used in sample preparation.

It has been reported recently that sample fractionation by SDS-PAGE is associated only with a moderate decrease of label-free quantitative measurement repeatability, while improving the depth of proteomic coverage (Gautier et al., 2012). However, we concluded based on our results that SDS-PAGE is not a suitable sample preparation method for label-free quantification and label-free results presented from here on are based on the shotgun experiments at a growth rate of 0.1 h^{-1}.

5.3.2 *Comparison of different label-free quantification methods (Publication III)*

In Publication III, three label-free proteome quantification methods — APEX, emPAI and iBAQ were compared in order to measure proteome-wide protein concentrations in the cells. All methods were applied to a shotgun sample from *E. coli* chemostat culture at growth rate 0.1 h^{-1}. Label-free quantification was made on an absolute scale, meaning that concentrations were calculated for each protein as a number of molecules per cell.

Absolute protein abundance in cells can be calculated by normalizing the contribution of individual proteins with the total protein mass in the cell (see for example (Nagaraj et al., 2011)). This is done using a quantitative measure, such as the sum of mass spectrometry responses of peptides used to identify each protein. However, this method is dependent on the measured total protein amount and on the number of identified proteins. Another approach would be to perform absolute quantification by spiking the sample with a known quantity of intact proteins and estimating protein concentration based on linear relationship of the mass spectrometry response and concentration of standard proteins.

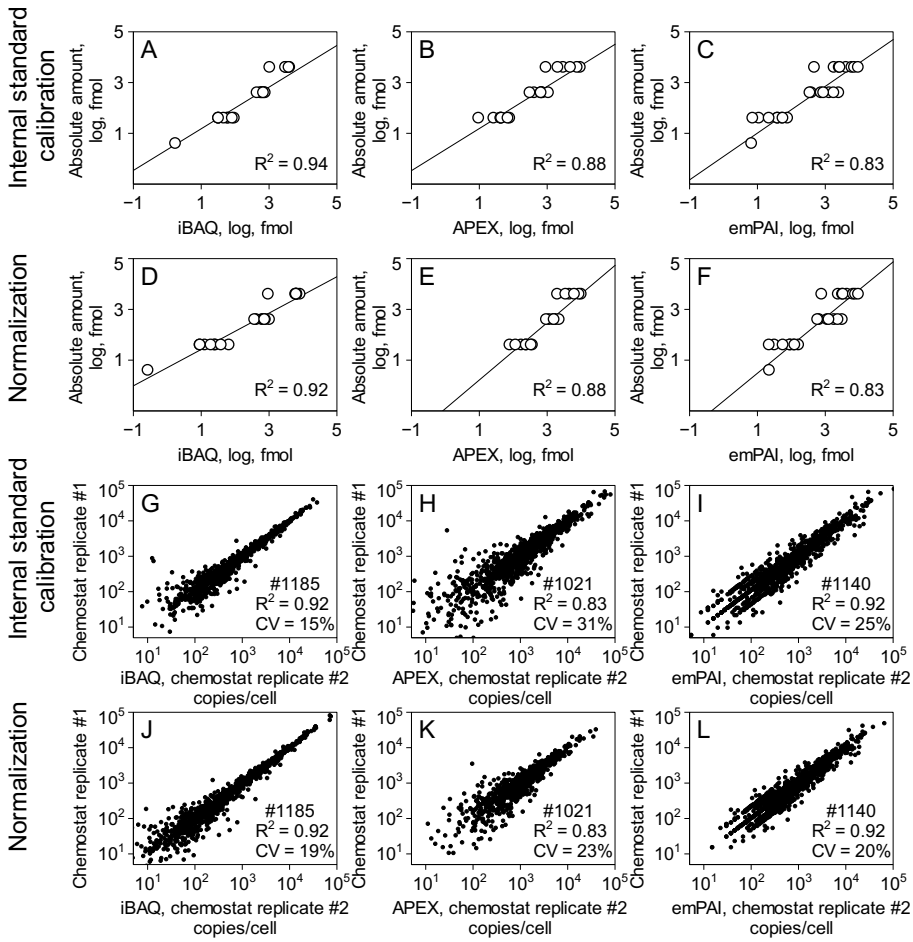


Figure 9 – The effect of internal standard calibration or normalization on protein concetration, calculated by label-free quantification methods. **A-C**) UPS2 protein abundances calculated by the internal standard calibration. **D-F**) UPS2 protein abundances calculated by the normalization method. **G-I**) Correlation between biological replicates calculated by the internal standard calibration. **J-L**) Correlation between biological replicates calculated by the normalization method.

In order to investigate the effect of internal standard addition on the performance of label-free quantification methods, Universal Proteomics Standard (UPS2, Sigma Aldrich) was used. UPS2 is a mixture of 48 precisely quantified human proteins with a dynamic concentrations range spanning five orders of magnitude. Internal standard addition enabled us to evaluate the magnitude of absolute protein abundances. The sum of all proteins in a cell, according to iBAQ and emPAI, was 8 and 5% less than the value derived from the Lowry total protein analysis. This very small difference between the total protein amount measured by the colorimetric assay and label-free quantitative proteomics meth-

ods indicates a high confidence of calculated protein abundances. Interestingly, the APEX method overestimates total protein concentration 1.5 times compared to Lowry, iBAQ and emPAI methods.

Comparison of standard protein abundances, calculated either by normalization or linear regression, revealed no difference in the squared Pearson's correlations for spectral counting methods APEX and emPAI (0.88 and 0.83, respectively) (Figure 9B, C, E, F). Correlation improved from 0.87 to 0.94 for iBAQ if the linear regression based on standard proteins was used instead of the normalization approach (Figure 9A compared to Figure 9D). Therefore, we decided to quantify protein abundance using the most appropriate approach for each method. Normalization was applied for APEX and emPAI, and internal standard calibration was used for iBAQ.

High correlation between biological replicates for all three absolute quantification methods was observed (Figure 9G, K and L). However, iBAQ outperformed the others: Pearson's squared correlation for logarithmized abundances 0.99 versus 0.92 for APEX and 0.89 for emPAI were measured (Figure 9G, K and L). High correlation between biological replicates can be explained by the highly reproducible continuous cultivation system and also due to minimized sample preparation by shotgun proteomics experiment.

Ribosomes are one of the largest protein complexes working in the cell and ribosomal proteins are expected to be expressed in equal copy numbers, however, we could not find agreement with the theoretical 1:1 stoichiometry. We identified and quantified 53 of 54 annotated ribosomal proteins and found that their absolute abundances span over one order of magnitude with the intensity based absolute quantification (iBAQ) method and over two orders of magnitude with spectral counting methods. Median ribosomal abundances were found to be 7,063, 3,987 and 5,219 copies per cell with CVs of 35%, 85% and 98% for iBAQ, APEX and emPAI, respectively (Publication III, Figure 4C). This significant difference of ribosomal protein abundances between different label-free methods showed that quantification methods must be chosen with great care. In the literature, spectral counting has resulted in higher variations of ribosomal proteins than peak area measurement, probably due to a saturation effect in spectral counting for such high abundant proteins (Ishihama et al., 2008; Maier et al., 2011). Because ribosomal proteins are relatively short and have high lysine and arginine content they produce a lot of tryptic peptides compared to their length which can complicate label-free quantification of these proteins.

Reasonable correlation between protein abundances determined by different absolute label-free quantification methods was observed. The spectral counting method APEX versus peak area calculation method iBAQ resulted in $R^2 = 0.76$, and correlation between another spectral counting method emPAI and iBAQ was $R^2 = 0.81$. Correlation between two spectral counting methods emPAI and APEX was found to be $R^2 = 0.77$ (Publication III, Figure 1G-I).

Median protein copy numbers per cell at growth rate 0.1 h^{-1} were 457, 886 and 409 for iBAQ, APEX and emPAI, respectively. We found that the top 20% of proteins by abundance contributed 76%, 62% and 78% of total protein amount in the cell for iBAQ, APEX and emPAI, respectively. This is in accordance with the well-known understanding that a small fraction of proteins are of high abundance. This has been observed for example in studies of mammalian cells (Beck et al., 2011; Nagaraj et al., 2011), *Saccharomyces cerevisiae*

(Ghaemmaghami et al., 2003), *Leptospira interrogans* (Schmidt et al., 2011), and *Mycoplasma pneumoniae* (Maier et al., 2011).

5.3.3 *Proteome distribution (Publication II and Publication III)*

The *E. coli* K12 MG1655 genome is 4.64 million bp long (Blattner et al., 1997) and contains approximately 4303 ORFs (according to UniProtKB, 17 october, 2012). The theoretical 2D gel of the *E. coli* genome was formed by plotting all proteins according to their theoretical isoelectric points and molecular weights (Figure 10).

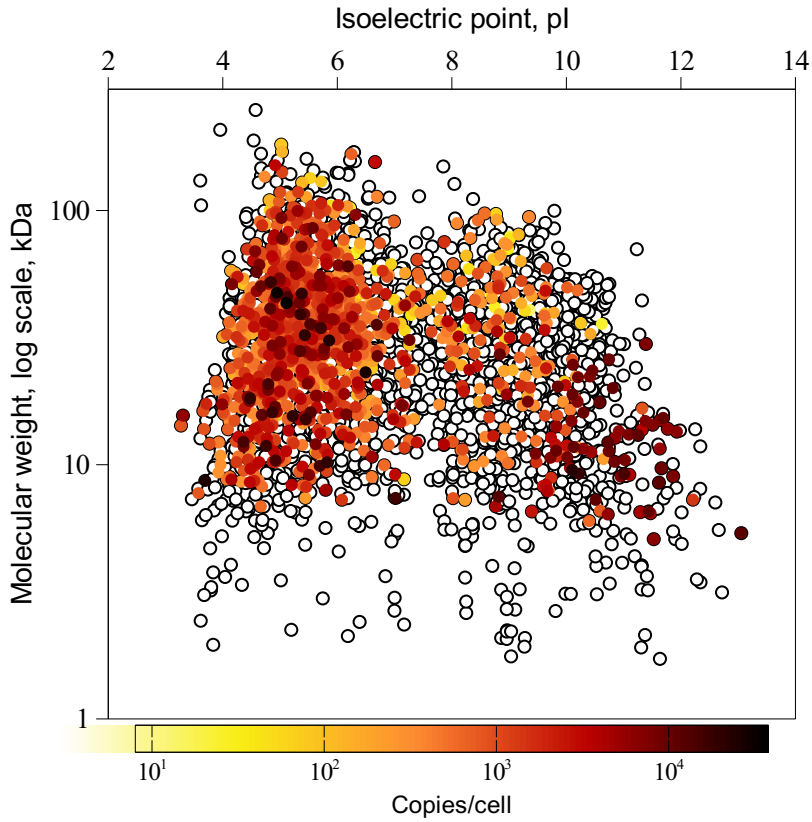


Figure 10 – Theoretical 2D map of all quantified *E. coli* proteins by the shotgun method. Empty spots represent proteins that remain unidentified in our analysis. Colour code represents quantitative information calculated by iBAQ method.

All proteins identified and quantified using the shotgun method were overlaid on the plot and colour coded according to the iBAQ absolute abundances. The theoretical 2D gel provides an overview of the molecular weight and isolectric point range of quantified proteins. Small proteins are missing from the analysis due to a limited number or lack of detectable tryptic peptides, making them difficult to identify by MS. However, our analysis

was not biased to large proteins, because 55% of the identified proteins had molecular weight over 30 kDa, well in accordance with the proportion of proteins (54%) larger than 30 kDa in the *E. coli* theoretical proteome. There is a clustering of identified proteins in the isoelectric point range of pH 4-7; this is expected because 65% of *E. coli* proteins are with isolectric point less than pH 7 (calculated based on theoretical isoelectric points of *E. coli* proteins).

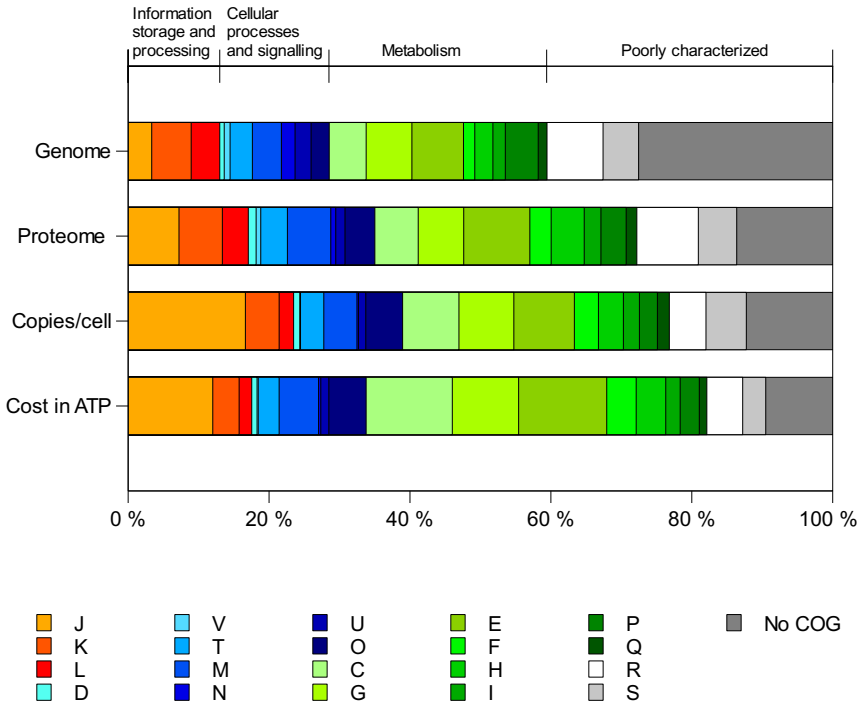


Figure 11 – Coverage and abundance levels of protein groups by COG classification. The number of genes, the number of identified proteins, the protein copies per cell and cost in ATP for protein synthesis. Functional classes are colour coded. J – translation, ribosomal structure and biogenesis; K – transcription; L – replication, recombination and repair; D – cell cycle control, cell division, chromosome partitioning; V – defence mechanisms; T – signal transduction mechanisms; M – cell wall/membrane/envelope biogenesis; N – cell motility; U – intracellular trafficking, secretion, and vesicular transport; O – posttranslational modification, protein turnover, chaperones; C – energy production and conversion; G – carbohydrate transport and metabolism; E – amino acid transport and metabolism; F – nucleotide transport and metabolism; H – coenzyme transport and metabolism; I – lipid transport and metabolism; P – inorganic ion transport and metabolism; Q – secondary metabolites biosynthesis, transport and catabolism; R – general function prediction only; S – function unknown; NO – no COG class.

The cellular roles of all *E. coli* K12 MG1655 4,303 protein coding genes and of all the identified proteins were predicted according to the classification of the Clusters of Orthologous Groups (COGs) database (Tatusov et al., 2003) maintained by the National Center for Biotechnology Information (NCBI). The data indicate a discrepancy between the number of protein coding genes and the number of identified proteins within COG classes (Figure 11, Genome and Proteome bars, respectively). The protein identifications are missing mostly for poorly characterized proteins. More than 40% of ORFs (28% of all the identified proteins) in *E. coli* are hypothetical, and do not have any COG class or are associated with a poorly characterized COG class.

By taking into account the protein copy numbers in the cell (Figure 11, Copies/cell bar) and the lengths of their polypeptide chains, together with the ATP cost of one amino acid polymerization reaction in the ribosome, the ATP cost for protein synthesis in the cell was calculated (Figure 11, Cost in ATP bar). This is a very rough calculation, because protein degradation is not considered. According to these calculations, *E. coli* invests a large fraction of cellular protein synthesis energy budget to the processes of translation, ribosomal structure and biogenesis (J); energy production and conversion (C), and amino acid transport and metabolism (E). In contrast, groups K, L, V, N and P have only a moderate impact on the abundance of proteins and the synthesis budget. The latter is partly caused by limitations of the analysis method — for example there is very little information regarding proteins from group N, which covers proteins involved in cell motility. Groups K and L embrace genes involved in transcription, replication, recombination and repair, which are of very low abundance and not costly for cells to synthesize. Group V, genes involved in defence mechanisms, are needed only in certain conditions and group P, inorganic ion transport and metabolism, are probably low abundant and low cost proteins. Poorly characterized proteins are generally at low abundance and do not demand excessive energetic costs.

5.3.4 *Protein dynamics within transcription units (Publication III)*

Proteins originating from the same transcription units should have similar absolute abundances, because they are synthesized from the same pool of mRNA species. The absolute abundance of proteins calculated in this study allowed us to quantitatively analyze the expression levels of proteins in transcription units. We divided transcriptional units into four groups based on their ratio to neighboring gene products: "up", "down", "others" and "no staircase" (see details in MATERIALS AND METHODS). "No staircase" regulation, where at least half of consecutive genes were not deferentially expressed, was found in only 9% of the transcription units (Figure 12). Most of the transcription units had significant differences between the genes. It has been previously demonstrated by computational methods that *E. coli* genes in transcription units are regulated equally (Laing et al., 2006). However, the previous analysis by Laing et al. (2006) only takes into account the average of all transcriptional units (Figure 12, All transcriptional units) and it eclipsed the staircase regulation which takes part in several modes. Our finding is also opposite to the findings in *M. pneumoniae* (Maier et al., 2011) and *L. interrogans* (Malmström et al., 2009) where staircase-like regulation on the protein level was not significant.

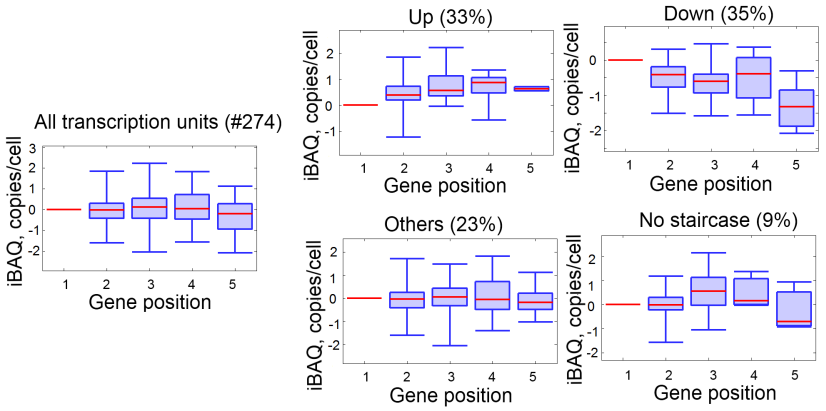


Figure 12 – Variation in protein dynamics within transcription units. Box plot diagrams for all ratios calculated for genes at position 1-5 in all quantified *E. coli* transcription units and grouped by "up", "down", "others" or "no staircase behavior".

Despite the differential regulation on transcriptional units the over all CV of quantified proteins was found to be 205%, which is more than three times higher than within the transcription units (60%) (Publication III, Supplementary Figure 7).

5.3.5 *Protein versus mRNA (Publication I, Publication II, Publication III)*

Relative proteomics expression data obtained using a low cost single MS instrument was compared with mRNA expression data measured by DNA microarray analysis for the same A-stat experiments (Nahku et al., 2010) (Publication I). High correlation between protein abundance and mRNA levels was not expected because slightly different growth rates (0.5/0.2 h^{-1} for proteome and 0.47/0.3 h^{-1} for transcriptome) and low-throughput imprecise proteome measurements were used. However, the Pearson's correlation 0.7 for 117 gene products was observed (when calculated based on log values then R^2 = 0.64), which indicated a good correlation (Publication I, Figure 3).

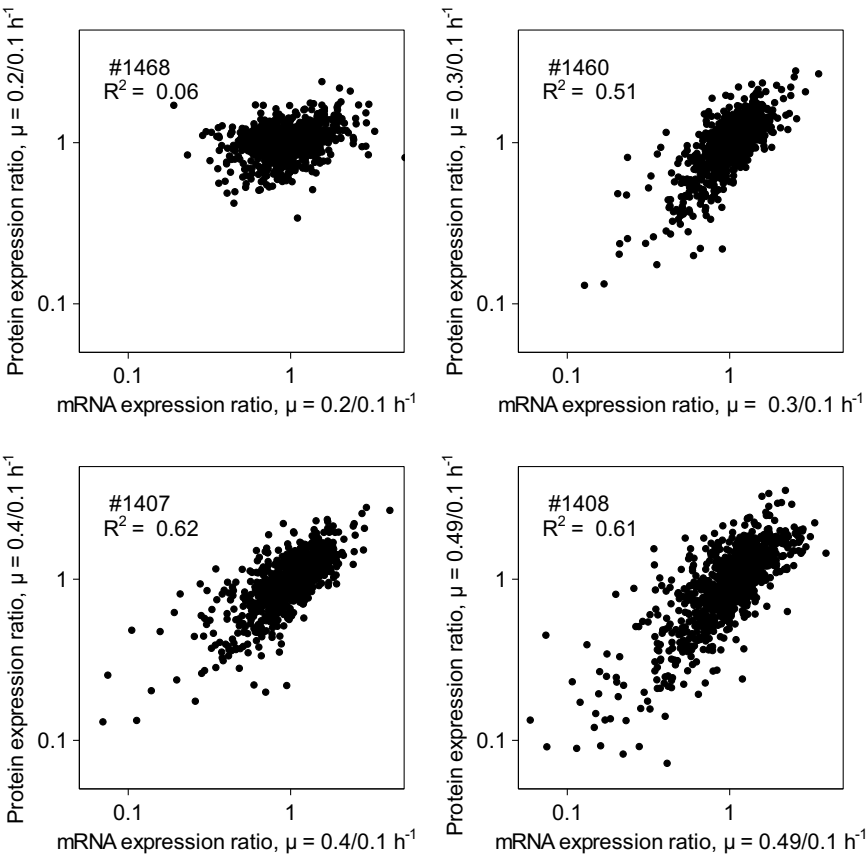


Figure 13 – Correlation of protein and mRNA expression ratios at different growth rates. Protein expression ratios obtained by using ^{15}N metabolically labeled reference, mRNA ratios obtained by using Agilent DNA microarray analysis. R^2 – Pearson's squared correlation coefficient; # – number of data points analyzed. All correlations were significant at p-value < 0.0001. Axes are at log scale.

Proteome expression, based on two biological replicates at five specific growth rates, were compared with mRNA levels in Publication II. The Pearson's squared correlation for more than 1,400 mRNA and protein pairs was R^2 = 0.51-0.62 for the specific growth rate range 0.3-0.48 h^{-1} compared with 0.1 h^{-1} (Figure 13). However, comparison of two low specific growth rate experiments, 0.2 h^{-1} and 0.1 h^{-1}, resulted in a very low Pearson's squared correlation (R^2 = 0.06). The latter was caused by two main reasons. First, no major metabolic changes occurred during this small change of conditions. Secondly, although there were small changes in mRNA levels, proteins levels were kept practically constant because it takes more time for protein levels to change due to the longer half-life of proteins (Figure 13) (Vogel and Marcotte, 2012; Maier et al., 2011).

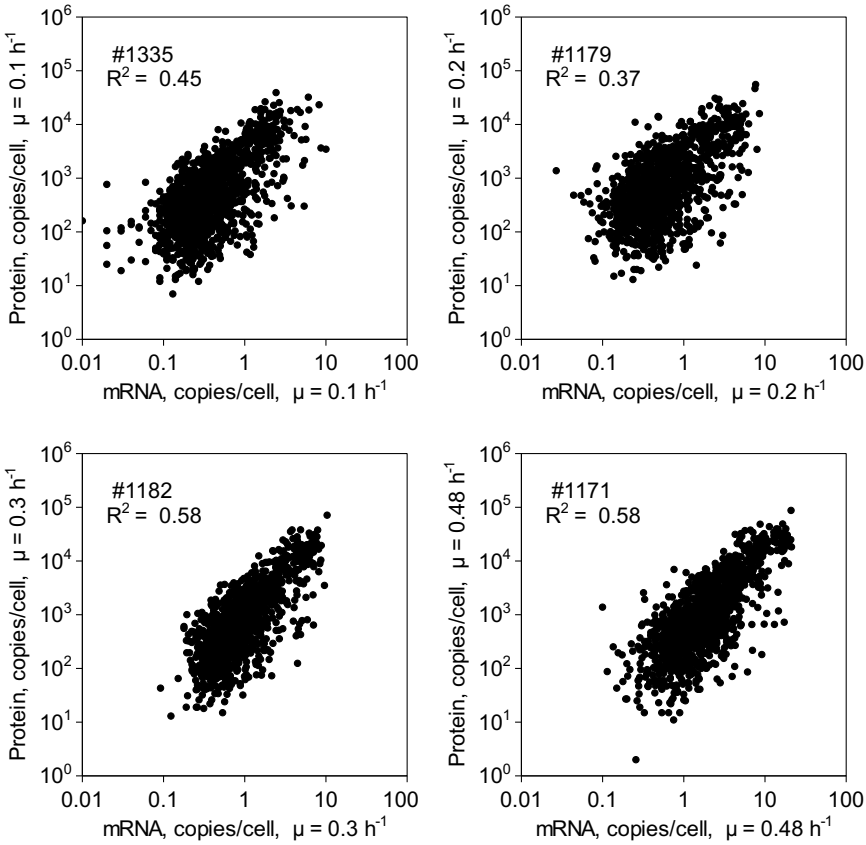


Figure 14 – Correlation of protein and mRNA absolute abundances at different growth rates. Protein abundances obtained by combining iBAQ values at growth rate 0.1 h^{-1} with relative ratios calculated by ^{15}N metabolically labeled reference. mRNA ratios obtained by using spot intensities from Agilent DNA microarray analysis. R^2 – Pearson's squared correlation coefficient; # – number of data points analyzed. All correlations are significant at p-value < 0.0001. Axes are at log scale. Correlation at growth rate 0.4 h^{-1} was 0.58 for 1,178 mRNA-proteins pairs.

By comparing the absolute abundances of mRNAs and proteins (Figure 14) squared Pearson's correlation 0.37-0.58 was found. Correlation was lower at low growth rates 0.1 and 0.2 h^{-1} which is also explained by longer protein half-lives than these of mRNA.

Correlations detected between mRNA and protein abundances in the current study are in accordance with the knowledge that ~40% of the variation in protein concentration can be explained by knowing mRNA abundances (Vogel and Marcotte, 2012; Maier et al., 2009; de Sousa Abreu et al., 2009).

5.3.6 *Apparent enzyme activity (Publication III)*

Apparent catalytic rates of enzymes (k_{cat}) were calculated per protein chain or subunits in order to estimate enzyme activities without *in vivo* assays in the cell at certain experimental conditions. Apparent k_{cat} was calculated for each enzyme at specific experimental condition as a ratio of specific flux and protein abundance designated for the respective flux (see MATERIALS AND METHODS for calculation details) (Figure 15). We found that biosynthetic enzymes (COG functional classes G, J, E, M, H, I, F) were working with ten times lower activity (median $< 10\ s^{-1}$) than energy generating enzymes (COG functional class C). High enzymatic activities of energy generating enzymes indicate a shortage of such genes, which may be a limiting factor for biomass or product formation rates.

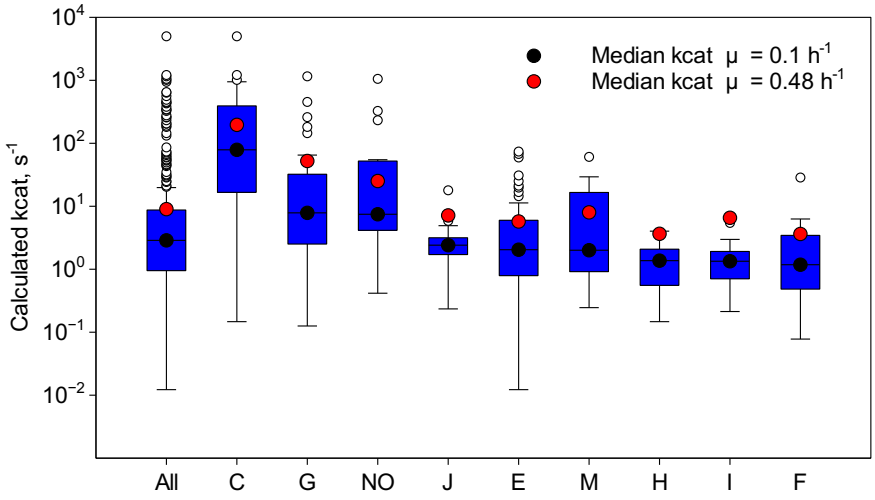


Figure 15 – Box plots showing the distribution of catalytic activities of enzymes at $\mu = 0.1\ h^{-1}$ divided into COGs. Horizontal bars represent 25^{th}, 50^{th} (median) and 75^{th} percentiles and whiskers represent 1.5 interquartile ranges. Outliers are plotted individually in open circles. Black dots represent median k_{cat} at slow and red dots at fast growth rate. All – median of all calculated enzymatic activities; C – energy production and conversion; G – carbohydrate transport and metabolism; NO – no COG class; J – Translation, ribosomal structure and biogenesis; E – amino acid transport and metabolism; M – cell wall/membrane/envelope biogenesis; H – coenzyme transport and metabolism; I – lipid transport and metabolism; F – nucleotide transport and metabolism.

The median k_{cat} value increased three times when comparing growth rate 0.48 h^{-1} and 0.1 h^{-1} indicating an increase in the metabolic capacity of *E. coli* at higher growth rates (Figure 15).

5.3.7 *Gene expression regulation*

As shown above, there was a positive correlation between levels of mRNAs and proteins, however, because the correlation was far from ideal, there was no transcriptional regulation for all gene products. In order to elucidate the regulation levels of gene expression, covariance analysis between protein per mRNA ratios and specific growth rates was used (details in MATERIALS AND METHODS) (Figure 16A). High uncertainty levels were found for almost half of the gene products due to the high error between biological replicate measurements for protein and mRNA abundances — no expression regulation could be revealed for those genes (ND, Figure 16). Almost half of the genes studied were found to be post-transcriptionally regulated (44%), which means that there were fewer proteins per mRNA translated with an increase in specific growth rate (Figure 16A). Genes were considered transcriptionally regulated if the protein to mRNA transcription rates were kept constant over the growth rates studied — this was the case for only 14% of the genes (Figure 16A). Translational regulation was determined only for minority of the genes, in this case translation rate was increased more than transcription rate with an increase in specific growth rate.

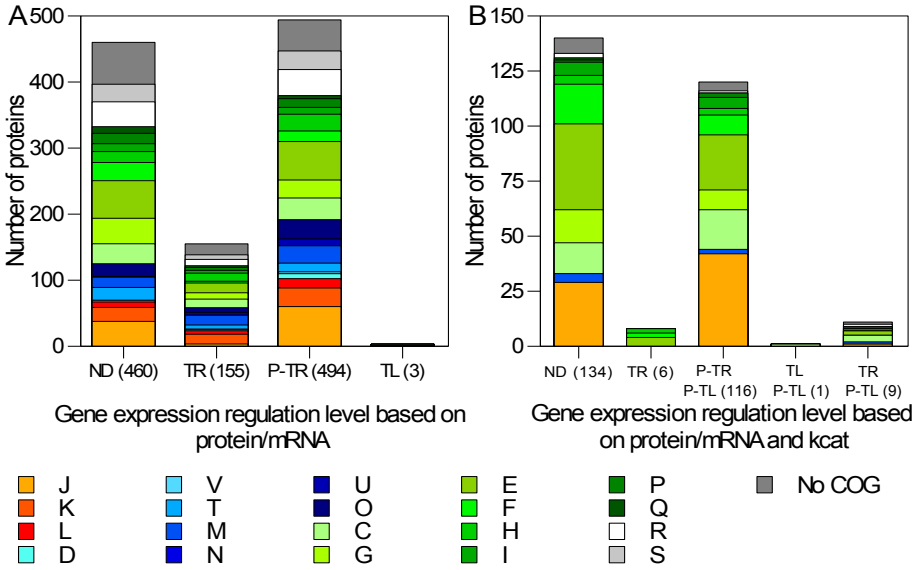


Figure 16 – Gene expression regulation. **A**) Based on covariance analysis of protein/mRNA ratio and specific growth rate; **B**) Based additionally to components in **A** also on covariance analysis of k_{cat} and specific growth rate. ND – no regulation analysis could be applied; TR – transcriptionally regulated; P-TR – post-transcriptionally regulated; TL – translationally; P-TL – post-translationally regulated. See Figure 11 for information of COG class nomenclature.

Post-translational regulation was added to the analysis by considering also covariance between k_{cat} and specific growth rate values (Figure 16B) — this was done for 266 proteins, for which a specific flux was calculated. Most of the post-translational regulations were combined with the post-transcriptional regulation (116 genes, Figure 16B), which means that when the protein to mRNA ratio decreased, the k_{cat} value increased with raising growth rate. The latter is probably regulated with PTMs. However PTMs were not analyzed in this study and this hypothesis should be tested in future studies.

Gene expression regulation was divided into COG functional classes and it was revealed that proteins involved in translation, ribosomal structure and biogenesis (group J) are mostly regulated at the post-transcriptional/post-translational level (Figure 16). Those proteins are mostly ribosomal proteins which are at high abundance and have low k_{cat} values. No other specific enrichment of gene expression in COG functional classes was detected, most probably due to high uncertainty of the analysis method.

Absolute quantification of proteomes combined with quantitative mRNA and metabolome analysis revealed that post-transcriptional gene regulation dominates in *E. coli* growth rate dependent studies. However, those calculations are a simplification because post-translational modifications, and protein and mRNA degradation rates were not measured. In addition, our data suffers from high uncertainty (for half of the mRNA-protein pairs).

SUMMARY

6

CONCLUSIONS

Four main conclusions result from this dissertation.

I A method for proteome characterization using a relatively low-cost single mass-spectrometry was developed and tested (Publication I). Peptide identifications obtained using a single mass analyzer were in a good correlation with results achieved with a tandem mass analyzer. The method is valuable for the detection of the most abundant proteins, especially if a limited numbers of high abundant proteins are of interest. The drawback of using a single mass analyzer for peptide identifying is the low selectivity, which is not comparable with tandem mass spectrometry. Because all of the ions are fragmented without any selection, mass spectra are often very complicated and our manual identification of peptides was very time consuming. To improve the performance of a single TOF-MS for peptide identification, a software routine should be developed to automate the peak picking process.

II The use of metabolically labeled culture as a spike-in standard is an advantageous technology for continuous culture experiments where the introduction of labeled media over the duration of the experiment is prohibitively expensive (Publication II).

- However, when using a spike-in standard culture, care should be taken that all of the interesting proteins are represented in the standard culture. The physiological state of the culture used for the labeled standard production and that obtained in the experiments must be similar.
- SDS-PAGE fractionation is a suitable sample preparation method for proteome quantification by metabolic labeling, where samples under study are mixed before fractionation.

III Absolute quantification of the proteome using three label-free quantification methods was validated (Publication III). The peak intensity (iBAQ) method was superior to spectral counting methods (APEX, emPAI) in terms of linearity of standard curves and reproducibility of biological replicates. In addition iBAQ provided the lowest variation among ribosomal protein abundances, which are expected to be present in equal amounts.

IV Absolute proteome quantification is essential for the comprehensive understanding of regulation mechanisms in the cell (Publication III). Absolute quantification allowed us to:

a) calculate the energetic burden put on cell energy generation system in order to synthesize proteins;

b) determine apparent enzyme activities at different growth rates of *E. coli*;

c) identify gene regulation mechanisms if combined with quantitative transcriptome and metabolome data.

BIBLIOGRAPHY

BIBLIOGRAPHY

Adamberg, K., Lahtvee, P. J., Valgepea, K., Abner, K., and Vilu, R. Quasi steady state growth of *Lactococcus lactis* in glucose-limited acceleration stat (A-stat) cultures. *Antonie van Leeuwenhoek*, 95(3):219–226, March 2009.

Aebersold, R. H., Teplow, D. B., Hood, L. E., and Kent, S. B. Electroblotting onto activated glass. High efficiency preparation of proteins from analytical sodium dodecyl sulfate-polyacrylamide gels for direct sequence analysis. *Journal of Biological Chemistry*, 261 (9):4229–4238, March 1986.

Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., and Kent, S. B. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis after *in situ* protease digestion on nitrocellulose. *Proceedings of the National Academy of Sciences of the United States of America*, 84(20):6970–6974, October 1987.

Ahlf, D. R., Compton, P. D., Tran, J. C., Early, B. P., Thomas, P. M., and Kelleher, N. L. Evaluation of the compact high-field Orbitrap for top-down proteomics of human cells. *Journal of Proteome Research*, 11(8):4308–4314, 2012.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. The shape and structure of proteins. In *Molecular Biology of the Cell*. Garland Science, New York, 4 edition, 2002.

Asara, J. M., Christofk, H. R., Freimark, L. M., and Cantley, L. C. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics*, 8(5):994–999, 2008.

Baliga, N. S., Pan, M., Goo, Y. A., Yi, E. C., Goodlett, D. R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W. V., and Hood, L. Coordinate regulation of energy transduction modules in *Halobacterium* sp. analyzed by a global systems approach. *Proceedings of the National Academy of Sciences*, 99(23):14913–14918, November 2002.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., and Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Analytical and Bioanalytical Chemistry*, 389(4):1017–1031, October 2007.

Bantscheff, M., Boesche, M., Eberhard, D., Matthieson, T., Sweetman, G., and Kuster, B. Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, 7(9):1702–1713, September 2008.

Beck, M., Schmidt, A., Malmstroem, J., Claassen, M., Ori, A., Szymborska, A., Herzog, F., Rinner, O., Ellenberg, J., and Aebersold, R. The quantitative proteome of a human cell line. *Molecular Systems Biology*, 7(1):1–8, November 2011.

Berrade, L., Garcia, A. E., and Camarero, J. A. Protein microarrays: Novel developments and applications. *Pharmaceutical Research*, 28(7):1480–1499, July 2011.

Biemann, K. Contributions of mass spectrometry to peptide and protein structure. *Biomedical and Environmental Mass Spectrometry*, 16(1-12):99–111, October 1988.

Blagoev, B., Ong, S.-E., Kratchmarova, I., and Mann, M. Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nature Biotechnology*, 22(9):1139–1145, 2004.

Blattner, F. R., Plunkett 3rd, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. The complete genome sequence of *Escherichia coli* k-12. *Science*, 277(5331):1453–1462, September 1997.

Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protocols*, 4(4):484–494, 2009.

Boisvert, F.-M., Ahmad, Y., Gierliński, M., Charrière, F., Lamont, D., Scott, M., Barton, G., and Lamond, A. I. A quantitative spatial proteomics analysis of proteome turnover in human cells. *Molecular & Cellular Proteomics*, 11(3):M111.011429, March 2012.

Braisted, J. C., Kuntumalla, S., Vogel, C., Marcotte, E. M., Rodrigues, A. R., Wang, R., Huang, S.-T., Ferlanti, E. S., Saeed, A. I., Fleischmann, R. D., Peterson, S. N., and Pieper, R. The APEX quantitative proteomics tool: Generating protein quantitation estimates from LC-MS/MS proteomics results. *BMC Bioinformatics*, 9:529, December 2008.

Brownridge, P., Holman, S. W., Gaskell, S. J., Grant, C. M., Harman, V. M., Hubbard, S. J., Lanthaler, K., Lawless, C., O'Cualain, R., Sims, P., Watkins, R., and Beynon, R. J. Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. *Proteomics*, 11(15):2957–2970, August 2011.

Brun, V., Dupuis, A., Adrait, A., Marcellin, M., Thomas, D., Court, M., Vandenesch, F., and Garin, J. Isotope-labeled protein standards: toward absolute quantitative proteomics. *Molecular & Cellular Proteomics*, 6(12):2139–2149, December 2007.

Brun, V., Masselon, C., Garin, J., and Dupuis, A. Isotope dilution strategies for absolute quantitative proteomics. *Journal of Proteomics*, 72(5):740–749, July 2009.

Buescher, J. M., Liebermeister, W., Jules, M., Uhr, M., Muntel, J., Botella, E., Hessling, B., Kleijn, R. J., Chat, L. L., Lecointe, F., Mäder, U., Nicolas, P., Piersma, S., Rügheimer, F., Becher, D., Bessieres, P., Bidnenko, E., Denham, E. L., Dervyn, E., Devine, K. M., Doherty, G., Drulhe, S., Felicori, L., Fogg, M. J., Goelzer, A., Hansen, A., Harwood, C. R., Hecker, M., Hubner, S., Hultschig, C., Jarmer, H., Klipp, E., Leduc, A., Lewis, P., Molina, F., Noirot, P., Peres, S., Pigeonneau, N., Pohl, S., Rasmussen, S., Rinn, B., Schaffer, M., Schnidder, J., Schwikowski, B., Dijl, J. M. V., Veiga, P., Walsh, S., Wilkinson, A. J., Stelling, J., Aymerich, S., and Sauer, U. Global network reorganization during dynamic adaptations of bacillus subtilis metabolism. *Science*, 335(6072):1099–1103, March 2012.

Canelas, A. B., Harrison, N., Fazio, A., Zhang, J., Pitkänen, J., Brink, J. v. d., Bakker, B. M., Bogner, L., Bouwman, J., Castrillo, J. I., Cankorur, A., Chumnanpuen, P., Daran-Lapujade, P., Dikicioglu, D., Eunen, K. v., Ewald, J. C., Heijnen, J. J., Kirdar, B., Mattila, I., Mensonides, F. I. C., Niebel, A., Penttilä, M., Pronk, J. T., Reuss, M., Salusjärvi, L., Sauer, U., Sherman, D., Siemann-Herzberg, M., Westerhoff, H., Winde, J. d., Petranovic, D., Oliver, S. G., Workman, C. T., Zamboni, N., and Nielsen, J. Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains. *Nature Communications*, 1:145, December 2010.

Cannon, J., Lohnes, K., Wynne, C., Wang, Y., Edwards, N., and Fenselau, C. High-throughput middle-down analysis using an Orbitrap. *Journal of Proteome Research*, 9(8):3886–3890, 2010.

Capriotti, A. L., Cavaliere, C., Foglia, P., Samperi, R., and Laganà, A. Intact protein separation by chromatographic and/or electrophoretic techniques for top-down proteomics. *Journal of Chromatography A*, 1218(49):8760–8776, December 2011.

Cargile, B. J., Talley, D. L., and Stephenson, J. L. Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. *Electrophoresis*, 25(6):936–945, March 2004.

Celis, J. E., Ratz, G. P., and Celis, A. Secreted proteins from normal and SV40 transformed human MRC-5 fibroblasts: toward establishing a database of human secreted proteins. *Leukemia: official journal of the Leukemia Society of America, Leukemia Research Fund, UK*, 1(10):707–717, October 1987.

Chelius, D. and Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research*, 1 (4):317–323, August 2002.

Chen, G., Gharib, T. G., Huang, C.-C., Taylor, J. M. G., Misek, D. E., Kardia, S. L. R., Giordano, T. J., Iannettoni, M. D., Orringer, M. B., Hanash, S. M., and Beer, D. G. Discordant protein and mRNA expression in lung adenocarcinomas. *Molecular & Cellular Proteomics*, 1(4):304–313, April 2002.

Choe, L., D'Ascenzo, M., Relkin, N. R., Pappin, D., Ross, P., Williamson, B., Guertin, S., Pribil, P., and Lee, K. H. 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics*, 7(20):3651–3660, October 2007.

Choi, H., Fermin, D., and Nesvizhskii, A. I. Significance analysis of spectral count data in label-free shotgun proteomics. *Molecular & Cellular Proteomics*, 7(12):2373–2385, December 2008.

Clamp, M., Fry, B., Kamal, M., Xie, X., Cuff, J., Lin, M. F., Kellis, M., Lindblad-Toh, K., and Lander, E. S. Distinguishing protein-coding and noncoding genes in the human genome. *Proceedings of the National Academy of Sciences*, 104(49):19428–19433, December 2007.

Collier, T. S., Sarkar, P., Franck, W. L., Rao, B. M., Dean, R. A., and Muddiman, D. C. Direct comparison of stable isotope labeling by amino acids in cell culture and spectral counting for quantitative proteomics. *Analytical Chemistry*, 82(20):8696–8702, October 2010.

Contiero, J., Beatty, C., Kumari, S., DeSanti, C. L., Strohl, W. R., and A, W. Effects of mutations in acetate metabolism on high-cell-density growth of *Escherichia coli*. *Journal of Industrial Microbiology and Biotechnology*, 24(6):421–430, 2000.

Costenoble, R., Picotti, P., Reiter, L., Stallmach, R., Heinemann, M., Sauer, U., and Aebersold, R. Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. *Molecular Systems Biology*, 7:464, February 2011.

Cox, J. and Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, 26(12):1367–1372, 2008.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. Andromeda: A peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research*, 10(4):1794–1805, 2011.

Craig, R. and Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics*, 20(9):1466–1467, June 2004.

Crick, F. Central dogma of molecular biology. *Nature*, 227(5258):561–563, August 1970.

Cristobal, A., Hennrich, M. L., Giansanti, P., Goerdayal, S. S., Heck, A. J. R., and Mohammed, S. In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis. *Analyst*, 137(15):3541–3548, July 2012.

Cunningham, Connell, J., Glish, G. L., and Burinsky, D. J. High amplitude short time excitation: a method to form and detect low mass product ions in a quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry*, 17(1): 81–84, January 2006.

de Godoy, L. M. F., Olsen, J. V., Cox, J., Nielsen, M. L., Hubner, N. C., Fröhlich, F., Walther, T. C., and Mann, M. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature*, 455(7217):1251–1254, September 2008.

de Groot, M. J. L., Daran-Lapujade, P., Breukelen, B. v., Knijnenburg, T. A., Hulster, E. A. F. d., Reinders, M. J. T., Pronk, J. T., Heck, A. J. R., and Slijper, M. Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. *Microbiology*, 153(11):3864–3878, November 2007.

de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., and Vogel, C. Global signatures of protein and mRNA expression levels. *Molecular BioSystems*, 5(12):1512–1526, November 2009.

Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D., and Alice, M. B. Molecular beams of macroions. *The Journal of Chemical Physics*, 49(5):2240–2249, September 1968.

Domon, B. and Aebersold, R. Mass spectrometry and protein analysis. *Science*, 312(5771): 212–217, April 2006.

Domon, B. and Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology*, 28(7):710–721, 2010.

Dressaire, C., Gitton, C., Loubière, P., Monnet, V., Queinnec, I., and Cocaign-Bousquet, M. Transcriptome and proteome exploration to model translation efficiency and protein stability in *Lactococcus lactis*. *PLoS Computatonal Biology*, 5(12):e1000606, December 2009.

Du, Y., Parks, B. A., Sohn, S., Kwast, K. E., and Kelleher, N. L. Top-down approaches for measuring expression ratios of intact yeast proteins using Fourier transform mass spectrometry. *Analytical Chemistry*, 78(3):686–694, 2005.

Durbin, K. R., Tran, J. C., Zamdborg, L., Sweet, S. M. M., Catherman, A. D., Lee, J. E., Li, M., Kellie, J. F., and Kelleher, N. L. Intact mass detection, interpretation, and visualization to automate top-down proteomics on a large scale. *Proteomics*, 10(20):3589–3597, October 2010.

Edman, P. A method for the determination of amino acid sequence in peptides. *Archives of Biochemistry*, 22(3):475, July 1949.

Edman, P. and Begg, G. A protein sequenator. *European Journal of Biochemistry*, 1(1): 80–91, March 1967.

Eidhammer, I., Flikka, K., Martens, L., and Mikalsen, S.-O. *Computational Methods for Mass Spectrometry Proteomics*. John Wiley & Sons, February 2008.

Eng, J., McCormack, A., and Yates, J. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *Journal of the American Society for Mass Spectrometry*, 5(11):976–989, 1994.

Engvall, E. and Perlmann, P. Enzyme-linked immunosorbent assay (ELISA) quantitative assay of immunoglobulin G. *Immunochemistry*, 8(9):871–874, September 1971.

Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science*, 246(4926):64–71, October 1989.

Fenselau, C. and Yao, X. $^{18}O_2$-labeling in quantitative proteomic strategies: a status report. *Journal of Proteome Research*, 8(5):2140–2143, May 2009.

Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., and Domon, B. Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, [Epub ahead of print], September 2012.

Gautier, V., Mouton-Barbosa, E., Bouyssié, D., Delcourt, N., Beau, M., Girard, J.-P., Cayrol, C., Burlet-Schiltz, O., Monsarrat, B., and Gonzalez de Peredo, A. Label-free quantification and shotgun analysis of complex proteomes by one-dimensional SDS-PAGE/NanoLC-MS: evaluation for the large scale analysis of inflammatory human endothelial cells. *Molecular and Cellular Proteomics*, 11(8):527–539, Aug 2012.

Geiger, T., Cox, J., and Mann, M. Proteomics on an orbitrap benchtop mass spectrometer using all-ion fragmentation. *Molecular & Cellular Proteomics*, 9(10):2252–2261, October 2010a.

Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., and Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods*, 7(5):383–385, May 2010b.

Geiger, T., Wehner, A., Schaab, C., Cox, J., and Mann, M. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Molecular and Cellular Proteomics*, 11(3):M111.014050, Mar 2012.

Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proceedings of the National Academy of Sciences of the United States of America*, 100(12):6940–6945, June 2003.

Getie-Kebtie, M., Lazarev, A., Eichelberger, M., and Alterman, M. Label-free mass spectrometry-based relative quantification of proteins separated by one-dimensional gel electrophoresis. *Analytical Biochemistry*, 409(2):202–212, February 2011.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R. W., Belle, A., Dephoure, N., O'Shea, E. K., and Weissman, J. S. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, October 2003.

Gibson, G. T. T., Mugo, S. M., and Oleschuk, R. D. Nanoelectrospray emitters: Trends and perspective. *Mass Spectrometry Reviews*, 28(6):918–936, 2009.

Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. 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):O111.016717, June 2012.

Gouw, J. W., Krijgsveld, J., and Heck, A. J. R. Quantitative proteomics by metabolic labeling of model organisms. *Molecular & Cellular Proteomics*, 9(1):11–24, January 2010.

Griffin, N. M., Yu, J., Long, F., Oh, P., Shore, S., Li, Y., Koziol, J. A., and Schnitzer, J. E. Label-free, normalized quantification of complex mass spectrometry data for proteomics analysis. *Nature Biotechnology*, 28(1):83–89, January 2010.

Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., and Aebersold, R. Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*. *Molecular & Cellular Proteomics*, 1(4):323–333, April 2002.

Grossmann, J., Roschitzki, B., Panse, C., Fortes, C., Barkow-Oesterreicher, S., Rutishauser, D., and Schlapbach, R. Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. *Journal of Proteomics*, 73(9): 1740–1746, August 2010.

Güell, M., Noort, V. v., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., Rode, M., Suyama, M., Schmidt, S., Gavin, A.-C., Bork, P., and Serrano, L. Transcriptome complexity in a genome-reduced bacterium. *Science*, 326(5957):1268–1271, November 2009.

Güell, M., Yus, E., Lluch-Senar, M., and Serrano, L. Bacterial transcriptomics: what is beyond the RNA horiz-ome? *Nature Reviews Microbiology*, 9(9):658–669, August 2011.

Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnology*, 17(10):994–999, October 1999a.

Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology*, 19(3):1720–1730, March 1999b.

Hager, J. W. A new linear ion trap mass spectrometer. *Rapid communications in mass spectrometry: RCM*, 16(6):512–526, 2002.

Hanke, S., Besir, H., Oesterhelt, D., and Mann, M. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. *Journal of Proteome Research*, 7(3):1118–1130, 2008.

Havliš, J. and Shevchenko, A. Absolute quantification of proteins in solutions and in polyacrylamide gels by mass spectrometry. *Analytical Chemistry*, 76(11):3029–3036, June 2004.

Hendrickson, E. L., Xia, Q., Wang, T., Leigh, J. A., and Hackett, M. Comparison of spectral counting and metabolic stable isotope labeling for use with quantitative microbial proteomics. *The Analyst*, 131(12):1335–1341, December 2006.

Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences*, 90(11):5011–5015, June 1993.

Hoehenwarter, W. and Wienkoop, S. Spectral counting robust on high mass accuracy mass spectrometers. *Rapid communications in mass spectrometry: RCM*, 24(24):3609–3614, December 2010.

Hoffmann, E. d. and Stroobant, V. *Mass Spectrometry: Principles and Applications*. Wiley-Interscience, 3 edition, November 2007.

Hörth, P., Miller, C. A., Preckel, T., and Wenz, C. Efficient fractionation and improved protein identification by peptide OFFGEL electrophoresis. *Molecular and Cellular Proteomics*, 5(10):1968–1974, October 2006.

Hoskisson, P. A. and Hobbs, G. Continuous culture — making a comeback? *Microbiology*, 151(10):3153–3159, October 2005.

Hsu, J.-L., Huang, S.-Y., Chow, N.-H., and Chen, S.-H. Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical Chemistry*, 75(24):6843–6852, December 2003.

Hubner, N. C., Ren, S., and Mann, M. Peptide separation with immobilized pI strips is an attractive alternative to in-gel protein digestion for proteome analysis. *Proteomics*, 8 (23-24):4862–4872, December 2008.

Iribarne, J. V. and Thomson, B. A. On the evaporation of small ions from charged droplets. *The Journal of Chemical Physics*, 64(6):2287–2294, March 1976.

Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., and Mann, M. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Molecular & Cellular Proteomics*, 4(9):1265–1272, September 2005.

Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J., and Frishman, D. Protein abundance profiling of the *Escherichia coli* cytosol. *BMC Genomics*, 9:102, February 2008.

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P. Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H., and Tomita, M. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, 316(5824):593–597, April 2007.

Iwasaki, M., Miwa, S., Ikegami, T., Tomita, M., Tanaka, N., and Ishihama, Y. One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the *Escherichia coli* proteome on a microarray scale. *Analytical Chemistry*, 82(7):2616–2620, 2010.

James, P. Protein identification in the post-genome era: the rapid rise of proteomics. *Quarterly Reviews of Biophysics*, 30(4):279–331, Nov 1997.

James, P., Quadroni, M., Carafoli, E., and Gonnet, G. Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications*, 195(1):58–64, August 1993.

Jayapal, K. P., Philp, R. J., Kok, Y.-J., Yap, M. G. S., Sherman, D. H., Griffin, T. J., and Hu, W.-S. Uncovering genes with divergent mRNA-protein dynamics in *Streptomyces coelicolor*. *PLoS ONE*, 3(5):e2097, May 2008.

Karas, M. and Hillenkamp, F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry*, 60(20):2299–2301, 1988.

Kelleher, N. L. A cell-based approach to the human proteome project. *Journal of the American Society for Mass Spectrometry*, 23(10):1617–1624, October 2012.

Keseler, I. M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muñiz-Rascado, L., Bonavides-Martinez, C., Paley, S., Krummenacker, M., Altman, T., Kaipa, P., Spaulding, A., Pacheco, J., Latendresse, M., Fulcher, C., Sarker, M., Shearer, A. G., Mackie, A., Paulsen, I., Gunsalus, R. P., and Karp, P. D. EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nucleic Acids Research*, 39(Database issue):D583–D590, January 2011.

Kim, B. H. and Gadd, G. M. *Bacterial Physiology and Metabolism*. Cambridge University Press, February 2008.

Köcher, T., Pichler, P., Schutzbier, M., Stingl, C., Kaul, A., Teucher, N., Hasenfuss, G., Penninger, J. M., and Mechtler, K. High precision quantitative proteomics using iTRAQ on an LTQ Orbitrap: a new mass spectrometric method combining the benefits of all. *Journal of Proteome Research*, 8(10):4743–4752, October 2009.

Kolkman, A., Daran-Lapujade, P., Fullaondo, A., Olsthoorn, M. M. A., Pronk, J. T., Slijper, M., and Heck, A. J. R. Proteome analysis of yeast response to various nutrient limitations. *Molecular Systems Biology*, 2(1):2006.0026, May 2006.

Krüger, M., Moser, M., Ussar, S., Thievessen, I., Luber, C. A., Forner, F., Schmidt, S., Zanivan, S., Fässler, R., and Mann, M. SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell*, 134(2):353–364, July 2008.

Kuntumalla, S., Braisted, J. C., Huang, S.-T., Parmar, P. P., Clark, D. J., Alami, H., Zhang, Q., Donohue-Rolfe, A., Tzipori, S., Fleischmann, R. D., Peterson, S. N., and Pieper, R. Comparison of two label-free global quantitation methods, APEX and 2D gel electrophoresis, applied to the *Shigella dysenteriae* proteome. *Proteome Science*, 7:22, June 2009.

Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227(5259):680–685, August 1970.

Lahtvee, P.-J., Adamberg, K., Arike, L., Nahku, R., Aller, K., and Vilu, R. Multi-omics approach to study the growth efficiency and amino acid metabolism in *Lactococcus lactis* at various specific growth rates. *Microbial Cell Factories*, 10:12, February 2011.

Laing, E., Mersinias, V., Smith, C., and Hubbard, S. Analysis of gene expression in operons of *Streptomyces coelicolor*. *Genome Biology*, 7(6):R46, June 2006.

Lamond, A. I., Uhlen, M., Horning, S., Makarov, A., Robinson, C. V., Serrano, L., Hartl, F. U., Baumeister, W., Werenskiold, A. K., Andersen, J. S., Vorm, O., Linial, M., Aebersold, R., and Mann, M. Advancing cell biology through proteomics in space and time (PROSPECTS). *Molecular & Cellular Proteomics*, 11(3):O112.017731, March 2012.

Lange, V., Picotti, P., Domon, B., and Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Molecular Systems Biology*, 4:222, October 2008.

Lanucara, F. and Eyers, C. E. Top-down mass spectrometry for the analysis of combinatorial post-translational modifications. *Mass Spectrometry Reviews*, [Epub ahead of print], 2012.

Larance, M., Bailly, A. P., Pourkarimi, E., Hay, R. T., Buchanan, G., Coulthurst, S., Xirodimas, D. P., Gartner, A., and Lamond, A. I. Stable isotope labeling with amino acids in nematodes. *Nature Methods*, 8(10):849–851, August 2011.

Le Roch, K. G., Johnson, J. R., Florens, L., Zhou, Y., Santrosyan, A., Grainger, M., Yan, S. F., Williamson, K. C., Holder, A. A., Carucci, D. J., Yates, J. R., and Winzeler, E. A. Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle. *Genome Research*, 14(11):2308–2318, November 2004.

Lee, J. E., Kellie, J. F., Tran, J. C., Tipton, J. D., Catherman, A. D., Thomas, H. M., Ahlf, D. R., Durbin, K. R., Vellaichamy, A., Ntai, I., Marshall, A. G., and Kelleher, N. L. A robust two-dimensional separation for top-down tandem mass spectrometry of the low-mass proteome. *Journal of the American Society for Mass Spectrometry*, 20(12): 2183–2191, December 2009.

Li, Z., Adams, R. M., Chourey, K., Hurst, G. B., Hettich, R. L., and Pan, C. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Velos. *Journal of Proteome Research*, 11(3):1582–1590, 2011.

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., and Darnell, J. *Molecular Cell Biology*. W. H. Freeman, New York, 4 edition, 2000.

Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology*, 25(1):117–124, 2007.

Ludwig, C., Claassen, M., Schmidt, A., and Aebersold, R. Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. *Molecular & Cellular Proteomics*, 11(3):M111.013987, March 2012.

Ma, B. and Johnson, R. *De novo* sequencing and homology searching. *Molecular & Cellular Proteomics*, 11(2):O111.014902, November 2011.

Maier, T., Güell, M., and Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Letters*, 583(24):3966–3973, December 2009.

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.-C., Aebersold, R., and Serrano, L. Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Molecular Systems Biology*, 7(1):1–12, July 2011.

Mallick, P. and Kuster, B. Proteomics: a pragmatic perspective. *Nature Biotechnology*, 28 (7):695–709, 2010.

Malmström, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W., and Aebersold, R. Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature*, 460(7256):762–765, August 2009.

Mann, M. and Wilm, M. Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry*, 66(24):4390–4399, December 1994.

Mann, M. Quantitative proteomics? *Nature Biotechnology*, 17(10):954–955, 1999.

Martin, S. F., Munagapati, V. S., Salvo-Chirnside, E., Kerr, L. E., and Le Bihan, T. Proteome turnover in the green alga *Ostreococcus tauri* by time course ^{15}N metabolic labeling mass spectrometry. *Journal of Proteome Research*, 11(1):476–486, January 2012.

Matic, I., Jaffray, E. G., Oxenham, S. K., Groves, M. J., Barratt, C. L. R., Tauro, S., Stanley-Wall, N. R., and Hay, R. T. Absolute SILAC-compatible expression strain allows sumo-2 copy number determination in clinical samples. *Journal of Proteome Research*, 10(10): 4869–4875, October 2011.

May, C., Brosseron, F., Chartowski, P., Schumbrutzki, C., Schoenebeck, B., and Marcus, K. Instruments and methods in proteomics. *Methods in Molecular Biology*, 696:3–26, 2011.

Michalski, A., Cox, J., and Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *Journal of Proteome Research*, 10(4):1785–1793, 2011a.

Michalski, A., Damoc, E., Hauschild, J., Lange, O., Wieghaus, A., Makarov, A., Nagaraj, N., Cox, J., Mann, M., and Horning, S. Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole orbitrap mass spectrometer. *Molecular & Cellular Proteomics*, 10(9):M111.011015, September 2011b.

Michalski, A., Damoc, E., Lange, O., Denisov, E., Nolting, D., Müller, M., Viner, R., Schwartz, J., Remes, P., Belford, M., Dunyach, J.-J., Cox, J., Horning, S., Mann, M., and Makarov, A. Ultra high resolution linear ion trap orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC-MS/MS and versatile peptide fragmentation modes. *Molecular & Cellular Proteomics : MCP*, 11(3):O111.013698, March 2012a.

Michalski, A., Neuhauser, N., Cox, J., and Mann, M. A systematic investigation into the nature of tryptic HCD spectra. *Journal of Proteome Research*, [Epub ahead of print], 2012b.

Mirgorodskaya, E., Braeuer, C., Fucini, P., Lehrach, H., and Gobom, J. Nanoflow liquid chromatography coupled to matrix-assisted laser desorption/ionization mass spectrometry: sample preparation, data analysis, and application to the analysis of complex peptide mixtures. *Proteomics*, 5(2):399–408, February 2005.

Mirzaei, H., McBee, J. K., Watts, J., and Aebersold, R. Comparative evaluation of current peptide production platforms used in absolute quantification in proteomics. *Molecular & Cellular Proteomics : MCP*, 7(4):813–823, April 2008.

Molina, H., Yang, Y., Ruch, T., Kim, J.-W., Mortensen, P., Otto, T., Nalli, A., Tang, Q.-Q., Lane, M. D., Chaerkady, R., and Pandey, A. Temporal profiling of the adipocyte proteome during differentiation using a five-plex SILAC based strategy. *Journal of Proteome Research*, 8(1):48–58, 2008.

Monod, J. Technique, theory and applications of continuous culture. *Annales de l'Institut Pasteur*, 79(4):390–410, October 1950.

Mørtz, E., O'Connor, P. B., Roepstorff, P., Kelleher, N. L., Wood, T. D., McLafferty, F. W., and Mann, M. Sequence tag identification of intact proteins by matching tanden mass spectral data against sequence data bases. *Proceedings of the National Academy of Sciences of the United States of America*, 93(16):8264–8267, August 1996.

Nagaraj, N., Wisniewski, J. R., Geiger, T., Cox, J., Kircher, M., Kelso, J., Pääbo, S., and Mann, M. Deep proteome and transcriptome mapping of a human cancer cell line. *Molecular Systems Biology*, 7(1):548, November 2011.

Nagaraj, N., Kulak, N. A., Cox, J., Neuhaus, N., Mayr, K., Hoerning, O., Vorm, O., and Mann, M. Systems-wide perturbation analysis with near complete coverage of the yeast proteome by single-shot UHPLC runs on a bench-top orbitrap. *Molecular & Cellular Proteomics*, 11(3):M111.013722, Mar 2012.

Nahku, R., Valgepea, K., Lahtvee, P.-J., Erm, S., Abner, K., Adamberg, K., and Vilu, R. Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures. *Journal of Biotechnology*, 145(1):60–65, January 2010.

Nakano, K., Rischke, M., Sato, S., and Märkl, H. Influence of acetic acid on the growth of *Escherichia coli* K12 during high-cell-density cultivation in a dialysis reactor. *Applied Microbiology and Biotechnology*, 48(5):597–601, Nov 1997.

Neidhardt, F. C. How microbial proteomics got started. *Proteomics*, 11(15):2943–2946, August 2011.

Nelson, C. J., Huttlin, E. L., Hegeman, A. D., Harms, A. C., and Sussman, M. R. Implications of ^{15}N-metabolic labeling for automated peptide identification in *Arabidopsis thaliana*. *Proteomics*, 7(8):1279–1292, April 2007.

Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry*, 75(17):4646–4658, 2003.

Nesvizhskii, A. I., Vitek, O., and Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods*, 4(10):787–797, October 2007.

Nie, L., Wu, G., and Zhang, W. Correlation between mRNA and protein abundance in desulfovibrio vulgaris: A multiple regression to identify sources of variations. *Biochemical and Biophysical Research Communications*, 339(2):603–610, January 2006.

Nilsen, T. W. and Graveley, B. R. Expansion of the eukaryotic proteome by alternative splicing. *Nature*, 463(7280):457–463, January 2010.

Novick, A. and Szilard, L. Description of the chemostat. *Science*, 112(2920):715–716, December 1950.

Oda, Y., Huang, K., Cross, F. R., Cowburn, D., and Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, 96(12):6591–6596, June 1999.

O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *The Journal of Biological Chemistry*, 250(10):4007–4021, May 1975.

Olsen, J. V., Godoy, L. M. F. d., Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., and Mann, M. Parts per million mass accuracy on an orbitrap mass spectrometer via lock mass injection into a C-trap. *Molecular & Cellular Proteomics*, 4(12):2010–2021, December 2005.

Olsen, J. V., Macek, B., Lange, O., Makarov, A., Horning, S., and Mann, M. Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods*, 4(9):709–712, September 2007.

Ong, S.-E. and Mann, M. Mass spectrometry-based proteomics turns quantitative. *Nature Chemical Biology*, 1(5):252–262, October 2005.

Ong, S.-E. and Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nature Protocols*, 1(6):2650–2660, 2007.

Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics*, 1(5): 376–386, May 2002.

Ono, M., Shitashige, M., Honda, K., Isobe, T., Kuwabara, H., Matsuzuki, H., Hirohashi, S., and Yamada, T. Label-free quantitative proteomics using large peptide data sets generated by nanoflow liquid chromatography and mass spectrometry. *Molecular & Cellular Proteomics*, 5(7):1338–1347, July 2006.

Paalme, T., Kahru, A., Elken, R., Vanatalu, K., Tiisma, K., and Raivo, V. The computer-controlled continuous culture of escherichia coli with smooth change of dilution rate (A-stat). *Journal of Microbiological Methods*, 24(2):145–153, December 1995.

Palmblad, M., Bindschedler, L. V., and Cramer, R. Quantitative proteomics using uniform ^{15}N-labeling, MASCOT, and the trans-proteomic pipeline. *Proteomics*, 7(19):3462–3469, October 2007.

Pan, S., Aebersold, R., Chen, R., Rush, J., Goodlett, D. R., McIntosh, M. W., Zhang, J., and Brentnall, T. A. Mass spectrometry based targeted protein quantification: methods and applications. *Journal of Proteome Research*, 8(2):787–797, February 2009.

Pappin, D. J., Hojrup, P., and Bleasby, A. J. Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology: CB*, 3(6):327–332, June 1993.

Patel, V. J., Thalassinos, K., Slade, S. E., Connolly, J. B., Crombie, A., Murrell, J. C., and Scrivens, J. H. A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *Journal of Proteome Research*, 8(7):3752–3759, 2009.

Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 20(18):3551–3567, December 1999.

Pichler, P., Köcher, T., Holzmann, J., Möhring, T., Ammerer, G., and Mechtler, K. Improved precision of iTRAQ and TMT quantification by an axial extraction field in an Orbitrap HCD cell. *Analytical Chemistry*, 83(4):1469–1474, February 2011.

Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B., and Aebersold, R. Full dynamic range proteome analysis of *Saccharomyces cerevisiae* by targeted proteomics. *Cell*, 138(4): 795–806, August 2009.

Picotti, P., Rinner, O., Stallmach, R., Dautel, F., Farrah, T., Domon, B., Wenschuh, H., and Aebersold, R. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. *Nature Methods*, 7(1):43–46, 2010.

Plumb, R., Castro-Perez, J., Granger, J., Beattie, I., Joncour, K., and Wright, A. Ultra-performance liquid chromatography coupled to quadrupole-orthogonal time-of-flight mass spectrometry. *Rapid communications in mass spectrometry: RCM*, 18(19):2331–2337, 2004.

Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., and Beynon, R. J. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nature Protocols*, 1(2):1029–1043, 2006.

Purvine, S., Eppel, J.-T., Yi, E. C., and Goodlett, D. R. Shotgun collision-induced dissociation of peptides using a time of flight mass analyzer. *Proteomics*, 3(6):847–850, June 2003.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., Berden, J. A., Brindle, K. M., Kell, D. B., Rowland, J. J., Westerhoff, H. V., Dam, K. v., and Oliver, S. G. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature Biotechnology*, 19(1):45–50, 2001.

Ramos, A. A., Yang, H., Rosen, L. E., and Yao, X. Tandem parallel fragmentation of peptides for mass spectrometry. *Analytical Chemistry*, 78(18):6391–6397, September 2006.

Rappsilber, J., Ryder, U., Lamond, A. I., and Mann, M. Large-scale proteomic analysis of the human spliceosome. *Genome Research*, 12(8):1231–1245, August 2002.

Rathsam, C., Eaton, R. E., Simpson, C. L., Browne, G. V., Valova, V. A., Harty, D. W. S., and Jacques, N. A. Two-dimensional fluorescence difference gel electrophoretic analysis of *Streptococcus* mutans biofilms. *Journal of Proteome Research*, 4(6):2161–2173, December 2005.

Ren, D., Pipes, G. D., Hambly, D., Bondarenko, P. V., Treuheit, M. J., and Gadgil, H. S. Top-down N-terminal sequencing of immunoglobulin subunits with electrospray ionization time of flight mass spectrometry. *Analytical Biochemistry*, 384(1):42–48, January 2009.

Renart, J., Reiser, J., and Stark, G. R. Transfer of proteins from gels to diazobenzyloxymethyl-paper and detection with antisera: a method for studying antibody specificity and antigen structure. *Proceedings of the National Academy of Sciences of the United States of America*, 76(7):3116–3120, July 1979.

Righetti, P. G. *Isoelectric Focusing: Theory, Methodology and Application: Theory, Methodology and Application*, volume 2. Elsevier, January 1983.

Riley, M., Abe, T., Arnaud, M. B., Berlyn, M. K., Blattner, F. R., Chaudhuri, R. R., Glasner, J. D., Horiuchi, T., Keseler, I. M., Kosuge, T., Mori, H., Perna, N. T., Plunkett, G., Rudd, K. E., Serres, M. H., Thomas, G. H., Thomson, N. R., Wishart, D., and Wanner, B. L. *Escherichia coli* K-12: a cooperatively developed annotation snapshot – 2005. *Nucleic Acids Research*, 34(1):1–9, 2006.

Roepstorff, P. and Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomedical Mass Spectrometry*, 11(11):601, November 1984.

Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D. J. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics*, 3(12):1154–1169, December 2004.

Ryu, S., Gallis, B., Goo, Y. A., Shaffer, S. A., Radulovic, D., and Goodlett, D. R. Comparison of a label-free quantitative proteomic method based on peptide ion current area to the isotope coded affinity tag method. *Cancer Informatics*, 6:243–255, April 2008.

Schaechter, M. From growth physiology to systems biology. *International Microbiology*, 9(3):157–161, Sep 2006.

Schmidt, A., Beck, M., Malmström, J., Lam, H., Claassen, M., Campbell, D., and Aebersold, R. Absolute quantification of microbial proteomes at different states by directed mass spectrometry. *Molecular Systems Biology*, 7(1):1–16, July 2011.

Schnölzer, M., Jedrzejewski, P., and Lehmann, W. D. Protease-catalyzed incorporation of ^{18}O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry. *Electrophoresis*, 17 (5):945–953, May 1996.

Schuchardt, S. and Sickmann, A. Protein identification using mass spectrometry: A method overview. In Baginsky, S. and Fernie, A. R., editors, *Plant Systems Biology*, volume 97 of *Experientia Supplementum*, pages 141–170. Birkhäuser Basel, 2007.

Schwanhäusser, B., Cox, J., Kirchner, M., Mann, M., and Selbach, M. Simple and accurate quantification of absolute protein abundance. (unpublished).

Schwanhäusser, B., Gossen, M., Dittmar, G., and Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics*, 9(1):205–209, January 2009.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. Global quantification of mammalian gene expression control. *Nature*, 473 (7347):337–342, May 2011.

Schwartz, J. C., Syka, J. P., and Quarmby, S. T. Improving the fundamentals of MS^n on 2D ion traps: New ion activation and isolation techniques. In *53rd ASMS Conference on Mass Spectrometry. San Antonio, Texas*, 2005.

Shen, Y., Zhao, R., Berger, S. J., Anderson, G. A., Rodriguez, N., and Smith, R. D. High-efficiency nanoscale liquid chromatography coupled on-line with mass spectrometry using nanoelectrospray ionization for proteomics. *Analytical Chemistry*, 74(16):4235–4249, 2002.

Siibak, T., Peil, L., Dönhöfer, A., Tats, A., Remm, M., Wilson, D. N., Tenson, T., and Remme, J. Antibiotic-induced ribosomal assembly defects result from changes in the synthesis of ribosomal proteins. *Molecular Microbiology*, 80(1):54–67, April 2011.

Silva, J. C., Denny, R., Dorschel, C., Gorenstein, M. V., Li, G.-Z., Richardson, K., Wall, D., and Geromanos, S. J. Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: A sweet tale. *Molecular & Cellular Proteomics*, 5(4):589–607, April 2006a.

Silva, J. C., Gorenstein, M. V., Li, G.-Z., Vissers, J. P. C., and Geromanos, S. J. Absolute quantification of proteins by $LCMS^E$: a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics*, 5(1):144–156, January 2006b.

Steen, H. and Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nature Reviews. Molecular Cell Biology*, 5(9):699–711, September 2004.

Stouthamer, A. H. A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie van Leeuwenhoek*, 39(3):545–565, 1973.

Sury, M. D., Chen, J.-X., and Selbach, M. The SILAC fly allows for accurate protein quantification *in vivo*. *Molecular & Cellular Proteomics*, 9(10):2173–2183, October 2010.

Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J., and Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26): 9528–9533, June 2004.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., Yoshida, T., and Matsuo, T. Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. *Rapid communications in mass spectrometry: RCM*, 2(8):151–153, 1988.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X. S. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, July 2010.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Smirnov, S., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, 4:41, September 2003.

Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., and Mann, M. Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. *Molecular & Cellular Proteomics*, 10(8):M110.003699, August 2011.

Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical Chemistry*, 75(8):1895–1904, 2003.

Tolonen, A. C., Haas, W., Chilaka, A. C., Aach, J., Gygi, S. P., and Church, G. M. Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Molecular Systems Biology*, 7:461, January 2011.

Tran, J. C. and Doucette, A. A. Gel-eluted liquid fraction entrapment electrophoresis: an electrophoretic method for broad molecular weight range proteome separation. *Analytical Chemistry*, 80(5):1568–1573, March 2008a.

Tran, J. C. and Doucette, A. A. Rapid and effective focusing in a carrier ampholyte solution isoelectric focusing system: a proteome prefractionation tool. *Journal of Proteome Research*, 7(4):1761–1766, April 2008b.

Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., Catherman, A. D., Durbin, K. R., Tipton, J. D., Vellaichamy, A., Kellie, J. F., Li, M., Wu, C., Sweet, S. M. M., Early, B. P., Siuti, N., LeDuc, R. D., Compton, P. D., Thomas, P. M., and Kelleher, N. L. Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*, 480(7376): 254–258, December 2011.

Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.-J., Arike, L., and Vilu, R. Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Systems Biology*, 4(1): 166, December 2010.

Valgepea, K., Adamberg, K., and Vilu, R. Decrease of energy spilling in *Escherichia coli* continuous cultures with rising specific growth rate and carbon wasting. *BMC Systems Biology*, 5:106, July 2011.

van Dongen, W. D., van Wijk, J. I., Green, B. N., Heerma, W., and Haverkamp, J. Comparison between collision induced dissociation of electrosprayed protonated peptides in the up-front source region and in a low-energy collision cell. *Rapid communications in mass spectrometry: RCM*, 13(17):1712–1716, 1999.

VanBogelen, R. A., Hutton, M. E., and Neidhardt, F. C. Gene-protein database of *Escherichia coli* K-12: edition 3. *Electrophoresis*, 11(12):1131–1166, December 1990.

Venable, J. D., Dong, M.-Q., Wohlschlegel, J., Dillin, A., and Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods*, 1(1):39–45, October 2004.

Vogel, C. and Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 13(4):227–232, March 2012.

Waanders, L., Hanke, S., and Mann, M. Top-down quantitation and characterization of SILAC-labeled proteins. *Journal of the American Society for Mass Spectrometry*, 18(11): 2058–2064, 2007.

Wagner, Y., Sickmann, A., Meyer, H. E., and Daum, G. Multidimensional nano-HPLC for analysis of protein complexes. *Journal of the American Society for Mass Spectrometry*, 14(9):1003–1011, September 2003.

Walsh, C. T., Garneau-Tsodikova, S., and Gatto, Gregory J, J. Protein posttranslational modifications: the chemistry of proteome diversifications. *Angewandte Chemie (International ed. in English)*, 44(45):7342–7372, December 2005.

Walther, T. C. and Mann, M. Mass spectrometry-based proteomics in cell biology. *The Journal of Cell Biology*, 190(4):491–500, August 2010.

Washburn, M. P., Wolters, D., and Yates III, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19(3): 242–247, March 2001.

Washburn, M. P., Koller, A., Oshiro, G., Ulaszek, R. R., Plouffe, D., Deciu, C., Winzeler, E., and Yates, J. R. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 100(6):3107–3112, March 2003.

Whiteaker, J. R., Lin, C., Kennedy, J., Hou, L., Trute, M., Sokal, I., Yan, P., Schoenherr, R. M., Zhao, L., Voytovich, U. J., Kelly-Spratt, K. S., Krasnoselsky, A., Gafken, P. R., Hogan, J. M., Jones, L. A., Wang, P., Amon, L., Chodosh, L. A., Nelson, P. S., McIntosh, M. W., Kemp, C. J., and Paulovich, A. G. A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nature Biotechnology*, 29(7):625–634, June 2011.

Wiese, S., Reidegeld, K. A., Meyer, H. E., and Warscheid, B. Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics*, 7(3): 340–350, February 2007.

Wilkins, M. R., Pasquali, C., Appel, R. D., Ou, K., Golaz, O., Sanchez, J. C., Yan, J. X., Gooley, A. A., Hughes, G., Humphery-Smith, I., Williams, K. L., and Hochstrasser, D. F. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (Nature Publishing Company)*, 14(1): 61–65, January 1996a.

Wilkins, M. R., Sanchez, J. C., Gooley, A. A., Appel, R. D., Humphery-Smith, I., Hochstrasser, D. F., and Williams, K. L. Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnology and Genetic Engineering Reviews*, 13:19–50, 1996b.

Williams, J. D., Flanagan, M., Lopez, L., Fischer, S., and Miller, L. A. D. Using accurate mass electrospray ionization-time-of-flight mass spectrometry with in-source collision-induced dissociation to sequence peptide mixtures. *Journal of Chromatography A*, 1020(1):11–26, December 2003.

Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature*, 379(6564):466–469, February 1996.

Wu, C., Tran, J. C., Zamdborg, L., Durbin, K. R., Li, M., Ahlf, D. R., Early, B. P., Thomas, P. M., Sweedler, J. V., and Kelleher, N. L. A protease for 'middle-down' proteomics. *Nature Methods*, 9(8):822–824, August 2012.

Xie, C., Ye, M., Jiang, X., Jin, W., and Zou, H. Octadecylated silica monolith capillary column with integrated nanoelectrospray ionization emitter for highly efficient proteome analysis. *Molecular & Cellular Proteomics*, 5(3):454–461, March 2006.

Xu, F., Xu, Q., Dong, X., Guy, M., Guner, H., Hacker, T. A., and Ge, Y. Top-down high-resolution electron capture dissociation mass spectrometry for comprehensive characterization of post-translational modifications in Rhesus monkey cardiac troponin I. *International Journal of Mass Spectrometry*, 305(2–3):95–102, August 2011.

Yan, W. and Chen, S. S. Mass spectrometry-based quantitative proteomic profiling. *Briefings in Functional Genomics and Proteomics*, 4(1):27–38, May 2005.

Yates, J. R., Ruse, C. I., and Nakorchevsky, A. Proteomics by mass spectrometry: Approaches, advances, and applications. *Annual Review of Biomedical Engineering*, 11(1): 49–79, 2009.

Yates III, J. R., Speicher, S., Griffin, P. R., and Hunkapiller, T. Peptide mass maps: a highly informative approach to protein identification. *Analytical Biochemistry*, 214(2):397–408, November 1993.

Zhang, W., Li, F., and Nie, L. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology*, 156(2):287–301, February 2010.

Zhang, Z. and Shah, B. Characterization of variable regions of monoclonal antibodies by top-down mass spectrometry. *Analytical Chemistry*, 79(15):5723–5729, 2007.

Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. Electron capture dissociation of multiply charged protein cations. a nonergodic process. *Journal of the American Chemical Society*, 120(13):3265–3266, 1998.

Zybailov, B., Coleman, M. K., Florens, L., and Washburn, M. P. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Analytical Chemistry*, 77 (19):6218–6224, 2005.

Zybailov, B., Mosley, A. L., Sardiu, M. E., Coleman, M. K., Florens, L., and Washburn, M. P. Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *Journal of Proteome Research*, 5(9):2339–2347, September 2006.

CURRICULUM VITAE

CURRICULUM VITAE

1. Personal data

First Name:	Liisa
Surname:	Arike
Date of Birth:	January 24, 1983
Place of Birth:	Tallinn, Estonia

2. Contact information

E-mail:	liisa@tftak.eu
Contact phone:	+37255635885
Address:	Kadaka pst. 29, Tallinn 10912, Estonia

3. Education

2007 - ...	Tallinn University of Technology, Faculty of Chemicals and Materials Technology, PhD student
2005 - 2007	Tallinn University of Technology, MSc in food and biotechnology,
2001 - 2005	Tallinn University of Technology, BSc in food and biotechnology

4. Professional Employment

2010 - ...	University of Tartu, Faculty of Science and Technology; Manager of Proteomics Core Facility.
2007.09 - 2008.02	Institute Technologie de Quimica u Biologia (New University of Lisbon), in the Mass Spectrometry Laboratory – Internship.
2005 –	Competence Center of Food and Fermentation Technologies; Researcher.

5. Teaching and supervising

13.08-17.08.2012	Practical Course "Quantitative Proteomics 2012", University of Tartu, Institute of Technology, Estonia. Organization, teaching and supervising practical work in laboratory.
30.08-03.09.2010	Practical Course "Quantitative Proteomics 2010", University of Tartu, Institute of Technology, Estonia. Supervising practical work in laboratory.
2009	Supervising BSc thesis of Reet Tsupilo "Analysis of Gram-positive and Gram-negative bacteria membrane proteins"**.**
2009 Autumn	Lecture course "Chemistry and Material science" (6 EAP, Tallinn Technical University), assistant of the practical work in the laboratory.

ELULOOKIRJELDUS

1. Isikuandmed

Eesnimi:	Liisa
Perekonnanimi:	Arike
Sünniaeg:	January 24, 1983
Sünnikoht:	Tallinn, Estonia

2. Kontaktandmed

E-mail:	liisa@tftak.eu
Telefon:	+37255635885
Address:	Kadaka pst. 29, Tallinn 10912, Eesti

3. Hariduskäik

2007 - ...	Tallinna Tehnikaülikool, Keemia-ja materjalitehnoloogia teaduskond, PhD tudeng
2005 - 2007	Tallinna Tehnikaülikool, Keemia-ja materjalitehnoloogia teaduskond – MSc bio-ja toiduainetetehnoloogias
2001 - 2005	Tallinna Tehnikaülikool, Keemia-ja materjalitehnoloogia teaduskond – BSc bio-ja toiduainetetehnoloogias

4. Teenistuskäik

2010 - ...	Tartu ülikool, Loodus- ja tehnoloogiateaduskond, proteoomika tuumiklabori juhataja
2007.09 - 2008.02	Keemia ja bioloogia tehnoloogiainstituut, Uus Lissaboni Ülikool, Portugal, praktika mass spektromeetria laboris
2005 –	Toidu-ja Fermentatsioonitehnoloogia Arenduskeskus, teadlane

5. Õpetamine ja juhendamine

13.08-17.08.2012	Praktiline kursus “Kvantitatiivne proteoomika”, Tartu ülikool, Tehnoloogiainstituut, organiseerimine, õpetamine ja laboratoorse töö juhendamine.
30.08-03.09.2010	Praktiline kursus “Kvantitatiivne proteoomika”, Tartu ülikool, Tehnoloogiainstituut, laboratoorse töö juhendamine.
2009	Reet Tsupilo BSc töö “Gram-negatiivsete ja Gram-positiivsete bakterite membraanivalkude analüüs” juhendamine.
2009 Sügis	“Anorgaaniline keemia - praktikum” Tallinna Tehnikaülikool, laboratoorse töö juhendamine

APPENDICES

PUBLICATION I

Arike L, Nahku R, Borissova M, Adamberg K, Vilu R

Identification and relative quantification of proteins in *Escherichia coli* proteome by "up-front" collision-induced dissociation

European Journal of Mass Spectrometry, 16(2):227-235 (2010)

Received: 31 August 2009 ■ *Revised: 9 January 2010* ■ *Accepted: 22 January 2010* ■ *Publication: 2 March 2010*

EJMS EUROPEAN JOURNAL OF MASS SPECTROMETRY

Identification and relative quantification of proteins in *Escherichia coli* proteome by "up-front" collision-induced dissociation

Liisa Arike,[a,b] Ranno Nahku,[a,c] Maria Borissova,[c] Kaarel Adamberg[a,b] and Raivo Vilu[a,c]

[a]Competence Centre of Food and Fermentation Technologies, Akadeemia tee 15b,12618 Tallinn, Estonia. E-mail: liisa@tftak.eu

[b]Tallinn University of Technology, Chemicals and Materials Science, Ehitajate tee 5, 19086 Tallinn, Estonia

[c]Tallinn University of Technology, Department of Chemistry, Ehitajate tee 5, 19086 Tallinn, Estonia

A method for identifying and quantifying proteins with relatively low-cost orthogonal acceleration time-of-flight mass spectrometry (oa-ToF-MS) was tested. *Escherichia coli* (*E. coli*) K12 MG1655 cell lysate was separated by 1D gel-electrophoresis; fractions were digested and separated fast and reproducibly by ultra-performance liquid chromatography (UPLC). Peptides were identified using oa-ToF-MS to measure exact masses of parent ions and the fragment ions generated by up-front collision-induced dissociation. Fragmentation of all compounds was achieved by rapidly cycling between high- and low values of energy applied to ions. More than 100 proteins from *E. coli* K12 proteome were identified and relatively quantified. Results were found to correlate with transcriptome data determined by DNA microarrays.

Keywords: up-front CID, ESI-ToF, 1D-LC-MS, label-free relative quantification, *E. coli* K12 proteome

Introduction

Application of collision-induced dissociation (CID) for fragmenting molecules is necessary for sequence analysis of peptides using electrospray ionization.[1] In tandem mass spectrometry (MS/MS) instruments, ions of interest are selected by the first analyzer and then directed into the collision cell where they collide with neutral gas molecules (for example argon or nitrogen). The second mass analyzer records all the fragments resulting in MS/MS spectra.[2,3] As well as fragmentation occurring in the collision cell, fragmentation can also be induced in one of the high-pressure regions between the atmospheric pressure source and the mass spectrometer. This process is called "up-front" collision induced dissociation[4] and has been referred to by a variety of terms including "nozzle-skimmer CID" and "in-source CID".[5] The difference between "up-front" CID and data dependent MS/MS experiments is that there is no precursor ion selection and fragments can result from the dissociation of all precursor ions, including solvent and background species.[5] It has been demonstrated that fragmentation of single- and double charged peptide ions by "up-front" CID can yield ion spectra comparable to those obtained in the collision cell of a tandem quadrupole-ToF mass spectrometer.[1] As the equipment is not spending any time in MS/MS mode, the data collection in an "up-front" CID scheme is much faster. Hence, more information can be obtained and no co-eluting peptides will be missed in narrow chromatographic peaks.[6]

In order to perform "up-front" CID in a single analyzer, the analyzer must operate in two different operating modes. In full-scan mode all ions from the source are analyzed, allowing the recording of precursor ions. In fragmentation mode all

ISSN: 1469-0667
doi: 10.1255/ejms.1068

ions are subjected to a fragmentation step which produces daughter ions.

The aim of this work was to test a low-cost alternative method for performing proteomic experiments. The fragmentation mode used in this paper has been described previously by Ren *et al.*[7] Briefly, in an oa-ToF-MS instrument, gas-phase ions from the source are transferred to the ToF mass analyzer by ion guiding tunnels, which consist of ion guide 1 and ion guide 2. Fragmentation occurs in the region between the ion guide 1 and the aperture separating the guides. In the ion guide 1 region, the ions gain significant energy as they accelerate. Collision of these ions with neutral molecules of the residual atmospheric gas leads to their dissociation.[7] A rapid change of voltage was applied to the aperture between the guides, and chromatograms for high- and low voltages were recorded in the same run. A similar switching approach was also used in a new technology called MS^e,[8,9] where the tandem quadrupole-ToF mass spectrometer is switched between low- and elevated collision energy.

Materials and methods

Preparation of peptide fractions

Escherichia coli (*E. coli*) K-12 MG1655 (Statens Serum Institute, Denmark) was cultivated in A-stat[10] culture, where specific growth rate was continuously increased according to the preset dilution rate. *E. coli* K-12 samples were collected at specific growth rates of 0.2 h^{-1} and 0.5 h^{-1}, centrifuged, washed once with phosphate buffer saline, centrifuged again and the pellet was suspended in sodium dodecyl sulfate (SDS) buffer [100 mM tris/HCl pH 7.5, 1% SDS, protease inhibitor cocktail (P8465, Sigma, USA)]. Cells were disrupted as a result of agitating the suspension with glass-beads at 4°C for 30 min. After centrifugation for 30 min at 4°C, the supernatant was collected and protein content was determined by 2D Quant kit (Amersham Biosciences, USA) and stored at –80°C prior to further analysis. Extracted proteins (60 µg) were separated on a 12% acrylamide sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel[11] (Protean II xi, BioRad, 20 × 20 cm). The gel was stained overnight with colloidal Coomassie, lanes were cut to 40 slices and slices were trypsin (Promega, Southampton, UK) digested.

Liquid chromatography-mass spectrometry

Peptides were analyzed with an oa-ToF instrument, LCT Premier, coupled with a UPLC instrument (Waters, UK). Five µL of peptide mixture (~0.25 µg) were loaded on a column (Waters ACQUITY UPLC BEH300 C18 1.7 µm, 2.1 mm × 100 mm) using gradient of 0.1% formic acid in ACN from 5–40% over 20 min in a flow rate of 0.15 mL min^{-1}. Data was collected with MassLynx 4.1 software (Waters, UK). The LCT Premier was run simultaneously in two modes, switching every second between low- and high energy (15 V and 55 V, respectively) in aperture between ion guides. Six fractions were also analyzed with a tandem quadrupole-ToF mass spectrometer, QStar Elite (Applied Biosystems, USA) in order to evaluate the identifications obtained with the LCT Premier.

Data analysis

Fragmented peptide spectra collected with LCT Premier were centroided in MassLynx 4.1 program and aligned manually with information about their precursor masses and charges and stored as DTA format files which were subsequently merged together. Proteins were identified by the Mascot search engine[12] using the NCBI database (17.10.2008). The search parameters were as follows; *E. coli* taxonomy, one missed trypsin cleavage, fixed modification: carbamidomethyl (C), variable modification: oxidation (M), 1.2 Da precursor mass tolerance and 0.6 Da fragment mass tolerance. A protein was considered positively identified if there were at least two peptides identified with a significant score.

The average MS signal response for the three most intense tryptic peptides was calculated for each identified protein.[8–13] For relative quantification, specific growth rate 0.5 h^{-1} was compared to 0.2 h^{-1}.

Results and discussion

Protein identification

A method for proteome characterization on relatively low-cost equipment was developed and tested. Simultaneous precursor ion measurement and peptide fragmentations were performed using up-front CID which yields to fragmentation of all the precursors entering the source. A similar approach has been applied to a mixture of known tryptic peptides.[6] In the current work, SDS-PAGE fractions from *E. coli* whole cell lysate were used. Gel fractions were digested and injected into liquid chromatography-mass spectrometry (LC-MS) where two different experiments were performed in the same run by rapidly changing between high- and low energies with concurrent collection of almost identical chromatograms. All the precursors which were entering the source were fragmented by changing fragmentation energies and both spectra were collected (Figure 1). Unlike a typical MS/MS experiment, where a precursor ion is isolated and analyzed individually, the precursor ions are fragmented without any selection in the current approach. Hence, the fragmented spectra can be much more complicated and lead to difficulties in fragment spectrum identification. The big drawback of the method is also the low automatization as peak lists are composed manually. Although the data produced is similar to MS^e data,[8,9] it is not yet possible to analyze it with a commercial software packages.

In the current study, more than 100 proteins (Table S1 in supplementary on-line material) were identified by manually created peak lists. A protein was considered identified with sufficient confidence when at least two peptides were identified. If the identification was based only on one sample but masses of the peptides were found with the same retention

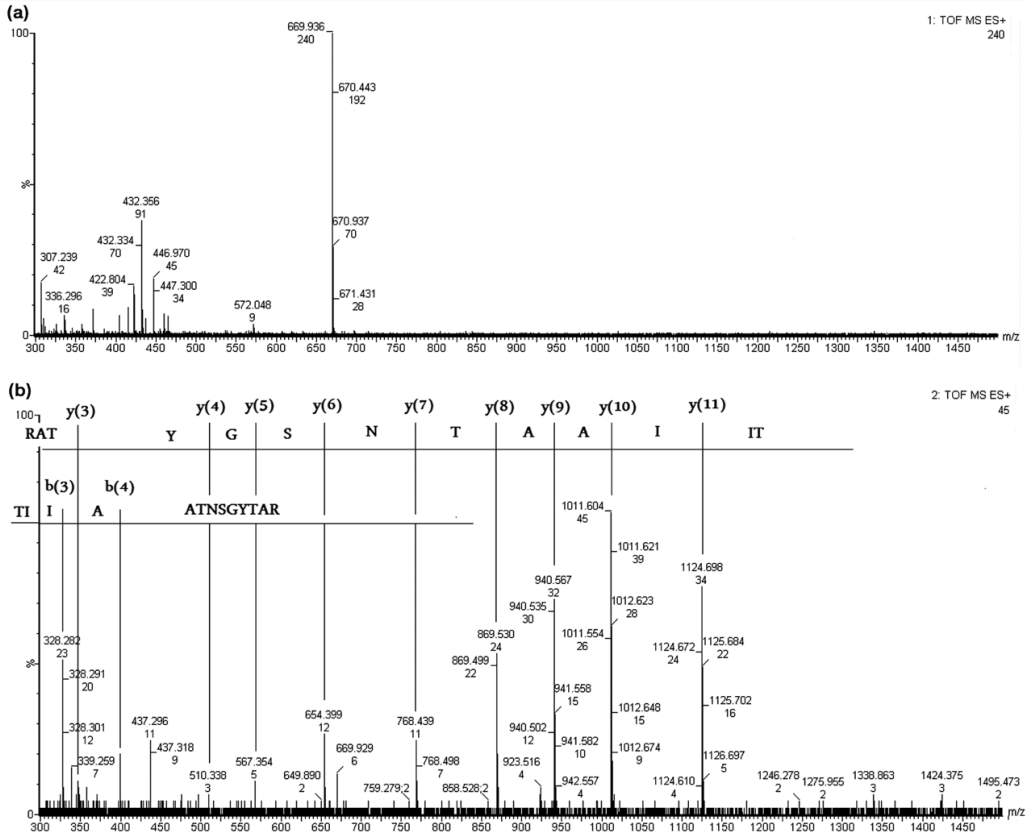


Figure 1. A two-step fragmentation method was employed for peptide TIIAATNSGYTAR. (a) First, the aperture 1 voltage was set to 15 V and peptide molecular masses were measured. (b) Next the voltage was rapidly switched to 55 V, leading to extensive dissociation which allowed identifying the partial sequence of the peptide.

time in both samples, the protein was considered to be present in both samples.

Proteins identified by the single ToF mass-spectrometer, LCT Premier, were compared with the results obtained by the tandem quadrupole-ToF mass analyzer, QStar Elite. Six different fractions were compared and all proteins which were observed with the LCT Premier were also found using QStar Elite (Figure 2). However, three times as many proteins were detected with the QStar Elite than with the LCT Premier [Figure 2(a)]. This remarkable advantage of QStar can at least be partly explained by the higher efficiency of automatic LC-MS/MS data analysis which was missing in the LCT Premier where peak lists were combined manually. It can only be assumed that if data handling with the LCT Premier had been automated, more peptides would have been detected with the oa-ToF as well. Despite the higher amount of proteins and peptides identified by MS/MS, there were 54 peptides which were identified only by oa-ToF [Figure 2(b)] (Table S2 in supplementary on-line materials).

Relative quantification

Silva, *et al.*[8] demonstrated that the average LC-MS signal response for the three most intensive tryptic peptides from a protein is correlating with the concentration of the protein. This method also allows for performing absolute quantification of proteins by spiking the cell lysate with known quantities of the proteins studied.

Quantification of peptides by LC-MS using peak integration is not often applied to proteins separated by gel electrophoresis. This is most likely due to the fact that sample processing during in-gel digestion and differences in the activities of trypsin in separate reaction vessels can possibly introduce undesired variations in the results.[13]

In the current study, SDS-PAGE separated proteins were quantified by comparing integrated peak areas and heights of the same peptides in different samples. It was assumed that the ratio between the average signal responses of the three most intensive peptides from one protein should correlate with the amount of the current protein. Latter hypothesis was

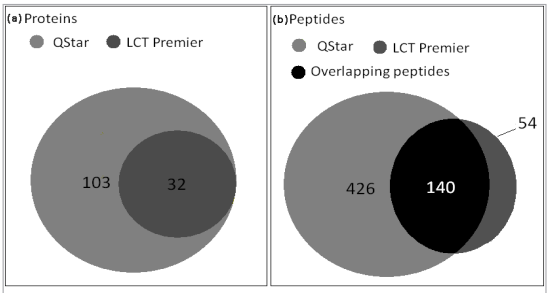


Figure 2. Evaluation of oa-ToF (LCT-Premier) results by comparison of six SDS gel fractions with tandem quadrupole-ToF mass spectrometer (QStar Elite). (a) All proteins observed with LCT Premier were also found with QStar Elite. (b) Almost four times more peptides were found with tandem quadrupole-ToF MS, however 54 unique peptides were identified only by oa-ToF.

positively tested with a standard protein bovine serum albumin (BSA). Different concentrations of BSA were loaded onto the same SDS gel as the samples and processed (data not shown). The calibration curve was found to be linear and no significant losses during the sample handling were observed.

Proteome changes—comparison with transcriptome changes

By now it has been understood that the detection of particular gene products in a microarray experiments does not confirm the presence or absence of the resulting protein product. This is due to the fact that the protein amount is being influenced by protein stability, translation rate, modulation of transcript levels by other proteins, post-translational modifications, half life etc.; therefore, mRNA levels cannot always be considered a predictor of the respective protein amount.[14] Bad correlation between mRNA and proteome levels has

Table 1. Differentially expressed (more than 1.5 fold up- or down-regulated) proteins (growth rate of 0.5 h^{-1} compared to 0.2 h^{-1}) and genes (growth rate of 0.47 h^{-1} compared to 0.3 h^{-1}) in *E. coli* A-stat culture.

Uniprot	Gene symbol	Protein name	Mass	Relative ratio of protein	Relative ratio of mRNA
P0AFG8	aceE	Pyruvate dehydrogenase	99948	0.56	1.16
P36682	acnB	Aconitate hydratase 1	93498	0.31	0.45
P27550	acs	Acetyl-coenzyme A synthetase	72094	0.08	0.05
P0AC41	sdhA	Succinate dehydrogenase flavoprotein subunit	64422	0.34	0.51
P0AG67	rpsA	30S ribosomal protein S1	61158	1.89	1.50
P0A7E5	pyrG	CTP synthase	60374	3.48	1.62
P0AC33	fumA	Fumarate hydratase class I, aerobic	60299	0.44	0.32
P08997	aceB	Malate synthase A	60274	0.47	0.45
P22259	pckA	Phosphoenolpyruvate carboxykinase [ATP]	59643	0.10	0.11
P02942	tsr	Methyl-accepting chemotaxis protein I	59443	0.64	0.81
P0A8N3	lysS	Lysyl-tRNA synthetase	57603	1.96	1.12
P0A6F5	groL	60 kDa chaperonin	57329	0.66	0.65
P0AFF6	nusA	Transcription elongation protein nusA	54871	1.88	1.28
P25553	aldA	Lactaldehyde dehydrogenase	52273	0.05	0.08
P00350	gnd	6-phosphogluconate dehydrogenase, decarboxylating	51481	2.67	1.43
P04949	fliC	Flagellin	51265	0.62	0.79
P39180	flu	Antigen 43 alpha chain	51000	0.57	0.63
P0A9P0	lpdA	Dihydrolipoyl dehydrogenase	50688	0.46	0.80

Table 1 (continued). Differentially expressed (more than 1.5 fold up- or down-regulated) proteins (growth rate of 0.5 h^{-1} compared to 0.2 h^{-1}) and genes (growth rate of 0.47 h^{-1} compared to 0.3 h^{-1}) in *E. coli* A-stat culture.

Uniprot	Gene symbol	Protein name	Mass	Relative ratio of protein	Relative ratio of mRNA
P0ABH7	gltA	Citrate synthase	48015	0.40	0.35
P0A9G6	aceA	Isocitrate lyase	47522	0.26	0.38
P0C8J8	gatZ	Putative tagatose 6-phosphate kinase gatZ	47109	0.32	0.24
P0A6P9	eno	Enolase	45655	2.21	1.22
P0A825	glyA	Serine hydroxymethyltransferase	45317	2.18	1.40
P0A836	sucC	Succinyl-CoA ligase [ADP-forming] subunit beta	41393	0.76	0.60
P0A799	pgk	Phosphoglycerate kinase	41276	3.11	1.08
P0A9Q9	asd	Aspartate-semialdehyde dehydrogenase	40018	2.41	1.29
P23721	serC	Phosphoserine aminotransferase	39783	3.57	1.41
P0AD96	livJ	Leucine/isoleucine/valine transporter subunit	39223	6.28	-
P02931	ompF	Outer membrane protein F	39333	0.48	0.75
P30178	ybiC	Uncharacterized oxidoreductase	38897	2.50	1.29
P0AB91	aroG	Phospho-2-dehydro-3-deoxyheptonate aldolase	38010	5.49	2.50
P0A9S3	gatD	Galactitol-1-phosphate 5-dehydrogenase	37390	0.07	0.13
P04391	argI	Ornithine carbamoyltransferase chain I	36907	2.40	1.47
P76316	dcyD	D-cysteine desulfhydrase	35153	1.53	1.06
P02931	ompF	Outer membrane protein F	39333	0.29	0.75
P61889	mdh	Malate dehydrogenase	32337	0.63	0.77
P0A6P1	tsf	Elongation factor Ts	30423	2.12	1.18
P0A9D8	dapD	Tetrahydrodipicolinate N-succinyltransferase	29892	2.20	1.11
P0AGE9	sucD	Succinyl-CoA ligase [ADP-forming] subunit alpha	29777	0.53	0.45
P0AEK4	fabI	NADH-dependent enoyl-ACP reductase	27864	1.75	1.09
P0C8J6	gatY	Tagatose-1,6-bisphosphate aldolase	30812	0.27	0.27
P28635	metQ	D-methionine-binding lipoprotein metQ	29432	0.62	0.87
P0A7L0	rplA	50S ribosomal protein L1	24730	1.67	1.33
P0AEM9	fliY	Cystine-binding periplasmic protein	29039	9.08	0.81
P39831	ydfG	NADP-dependent L-serine/L-allo-threonine dehydrogenase ydfG	27249	0.66	0.87
P07014	sdhB	Succinate dehydrogenase iron-sulfur subunit	26770	0.21	0.37

Table 1 (continued). Differentially expressed (more than 1.5 fold up- or down-regulated) proteins (growth rate of 0.5 h^{-1} compared to 0.2 h^{-1}) and genes (growth rate of 0.47 h^{-1} compared to 0.3 h^{-1}) in *E. coli* A-stat culture.

Uniprot	Gene symbol	Protein name	Mass	Relative ratio of protein	Relative ratio of mRNA
P69441	adk	Adenylate kinase	23586	3.39	1.48
P60438	rplC	50S ribosomal protein L3	22244	1.62	1.49
P0ACJ8	crp	Catabolite gene activator	23640	0.60	0.73
P0A8F0	upp	Uracil phosphoribosyltransferase	22533	2.90	1.54
P30126	leuD	3-isopropylmalate dehydratase small sub-unit	22487	1.98	1.10
P0A955	eda	KHG/KDPG aldolase	22284	3.19	1.19
P0AGD3	sodB	Superoxide dismutase [Fe]	21310	3.03	0.99
P0AE08	ahpC	Alkyl hydroperoxide reductase subunit C	20862	0.57	0.79
P62399	rplE	50S ribosomal protein L5	20302	2.41	1.16
P02359	rpsG	30S ribosomal protein S7	20019	2.42	0.92
P0AG55	rplF	50S ribosomal protein L6	18904	1.53	1.12
P0A862	tpx	Thiol peroxidase	17835	1.76	0.90
P0A7W1	rpsE	30S ribosomal protein S5	17603	2.13	1.15
P0A7F3	pyrI	Aspartate carbamoyltransferase regulatory chain	17121	3.80	1.61
P0A7J3	rplJ	50S ribosomal protein L10	17712	1.66	1.02
P0ADY7	rplP	50S ribosomal protein L16	15281	3.32	1.17
P02413	rplO	50S ribosomal protein L15	14980	1.61	1.17
P0ACF8	hns	DNA-binding protein H-NS	15540	2.39	0.76
P0A7J7	rplK	50S ribosomal protein L11	14875	4.77	1.67
P0A7W7	rpsH	30S ribosomal protein S8	14127	2.03	1.12
P0A7R9	rpsK	30S ribosomal protein S11	13845	2.52	1.09
P0ADY3	rplN	50S ribosomal protein L14	13541	2.62	0.94
P0A6F9	groS	10 kDa chaperonin	10387	1.57	0.87
P0A7K6	rplS	50S ribosomal protein L19	13133	2.14	1.56
P0C018	rplR	50S ribosomal protein L18	12770	3.61	1.13
P60624	rplX	50S ribosomal protein L24	11316	2.65	0.92
P0A7K2	rplL	50S ribosomal protein L7/L12	12295	1.74	1.14
P0A7R5	rpsJ	30S ribosomal protein S10	11736	2.51	1.33
P0A7T7	rpsR	30S ribosomal protein S18	8986	1.86	1.25

been reported by several groups, including for organisms such as *Plasmodium falciparum*,[15] *Saccharamyces cerevisiae*[16,17] and mouse.[18] However, it has also been reported that there is a good correlation between protein and transcript levels during exponential growth of *E. coli*.[19] Considering all the latter, it seems that data which monitors transcriptome and protein changes at the same time would give much more information about the physiological regulation of the studied microorganism.

Proteome data obtained in this study was compared with DNA microarray analysis results from the same experiments.[20] Relative changes in *E. coli* proteome were calculated for A-stat culture, comparing specific growth rates at $0.5\,h^{-1}$ to $0.2\,h^{-1}$. Relative changes in the transcriptome were found at the comparison of growth rates at $0.47\,h^{-1}$ to $0.3\,h^{-1}$. At higher growth rates ($> 0.34\,h^{-1}$), overflow metabolism was observed indicated by acetate production and lowered production yield of CO_2. Analyzed specific growth rate points for transcriptome and proteome comparisons can be taken as similar to those at lower growth rate ($< 0.34\,h^{-1}$) characteristics of cells (glucose, O_2 consumption and CO_2 production yields). However, at the growth rate of $0.5\,h^{-1}$, there was a minor amount of glucose (1 mM) in the growth environment which may induce some additional differences between mRNA and protein relative ratios.

Proteins were considered to be differentially expressed if there was at least a 1.5-fold change (relative ratio > 1.5 or < 0.67) (Table 1). In general, the changes in proteome and transcriptome were found to correlate by the Pearson coefficient of 0.7 (Figure 3), although there seems to be a trend towards stronger up-regulation of proteins. This can be caused by a higher specific growth rate range in the case of protein measurement.

Main changes at higher growth rates were found in acetate utilization and transport genes (*acs*, *yjcH* and *actP*), which were strongly repressed in transcriptome (> 10 times), indicating that acetate assimilation was decreased by this pathway. The down-regulation of Acs at the same level was also detected in proteome.

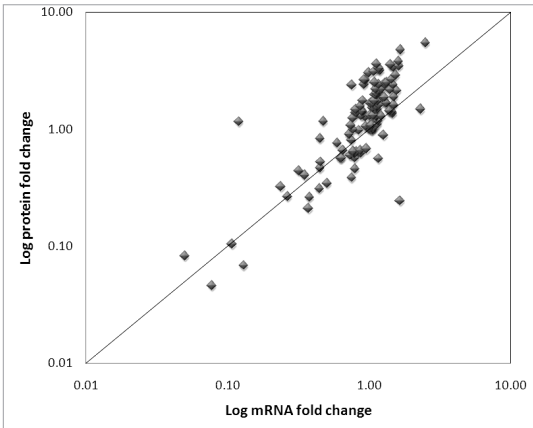


Figure 3. Protein and mRNA expression were found to correlate by Pearson coefficient of 0.7.

Decreased amounts of TCA cycle proteins (AcnB, GltA, FumA, SdhAB, SucCD) were observed at higher growth rates, similar to transcriptome levels (*acnB, gltA, fumAC, sdhABCD*, *sucABCD*). Also, the glycoxylate bypass proteins (AceAB) were down regulated at higher growth rate which is in concordance with the results obtained from transcriptome measurements.

Due to carbon catabolic repression at higher growth rate, down-regulation was observed for proteins responsible for the utilization and transport of substrates other than glucose: AldA, GatDYZ. The latter was confirmed by transcriptome measurements showing down regulation of genes *aldA, gatABCDYZ, lamB, malEFKMQ, manXZY, mglABC, rbsABCDK*.

As the growth rate increases, bacteria need to synthesize more cell building blocks such as DNA, RNA, proteins and ribosomes. At the higher growth rate, up-regulation was detected for proteins (ArgI, AroG, Asd, DapD, LeuD, SerC) and to a lesser amount for some genes (*argFH, aroGH, asd, serAC*) responsible for amino acid biosynthesis. Twenty six ribosomal proteins were found and all of them were slightly up-regulated by proteomic analysis. Interestingly, mRNA levels of ribosomal proteins were practically constant.

Conclusion

The method presented in this paper is an alternative method for performing proteomic experiments. It has been demonstrated that using a simple oa-ToF-MS instrument enables biologically important proteins to be identified and their relative amounts to be determined by comparing mass spectrometry peak intensities of the same peptides in different runs. Results obtained from single ToF mass analyzer identification were in a good correlation with identifications achieved with the tandem quadrupole-ToF mass analyzer although only the most abundant proteins were detected. The drawback of using a single ToF mass analyzer for peptide identification is the low selectivity which is not comparable with tandem quadrupole-ToF mass-spectrometry. As all the ions are fragmented without any selection, spectrums are often very complicated and manual identification of peptides is very time consuming. The performance of a single ToF MS could be improved with software that could do automatic peak picking. At the moment, such software does not exist to our knowledge. We believe that the current method could be applied if comparison of high abundance proteins (for example over-expressed proteins) is needed and access to tandem mass spectrometry is limited.

Comparison of proteome and transcriptome data showed that different data only complement each other and further comparative studies would be preferable to get a better picture of the correlation between proteome and transcriptome.

Acknowledgments

The financial support for this research was provided by the Enterprise Estonia project EU22704 and Ministry of Education, Estonia, through the grant SF0140090s08.

References

1. W.D. van Dongen, J.I.T. van Wijk, B.N. Green, W. Heerma and J. Haverkamp, "Comparison between collision induced dissociation of electrosprayed protonated peptides in the up-front source region and in a low-energy collision cell", *Rapid Commun. Mass Spectrom.* **13,** 1712–1716 (1999). doi: 110.1002/(SICI)1097-0231(19990915)13:17<1712::AID-RCM703>3.0.CO;2-8
2. I. Eidhammer, K. Flikka, L. Martens and S. Mikalsen, "Fragmentation models", in *Computational Methods for Mass Spectrometry Proteomics*. Wiley-Interscience, New York, USA, p. 284 (2008).
3. J.R. Chapman, "Ionization methods and instrumentation", in *Protein and Peptide Analysis by Mass Spectrometry*, Ed by J. Chapman. Human Press, Totowa, New Jersey, p. 350 (1996). doi: 10.1385/0-89603-345-7:9
4. A. Bruins, "ESI source design and dynamic range considerations", in *Electrospray Ionization Mass Spectrometry*, Ed by R.B. Cole. John Wiley & Sons, New York, USA, p. 577 (1997).
5. R.B. Cody, "Electrospray ionization mass spectrometry: history, theory and instrumentation", in *Applied Electrospray Mass Spectrometry*, Ed by B.N. Pramanik, A.K. Ganguly and M.L. Gross. CRC Press, New York, USA, p. 434 (2002).
6. J.D. Williams, M. Flanagan, L. Lopez, S. Fischer and L.A.D. Miller, "Using accurate mass electrospray ionization-time-of-flight mass spectrometry with in-source collision-induced dissociation to sequence peptide mixtures", *J. Chromatogr. A* **1020,** 11–26 (2003). doi: 10.1016/j.chroma.2003.07.019
7. D. Ren, G.D. Pipes, D. Hambly, P.V. Bondarenko, M.J. Treuheit and H.S. Gadgil, "Top-down N-terminal sequencing of immunoglobulin subunits with electrospray ionization time of flight mass spectrometry", *Anal. Biochem.* **384,** 42–48 (2009). doi: 10.1016/j.ab.2008.09.026
8. J.C. Silva, M.V. Gorenstein, G. Li, J.P.C. Vissers and S.J. Geromanos, "Absolute quantification of proteins by LC-MS^{e}: a virtue of parallel MS acquisition", *Mol. Cell. Proteomics* **5,** 144-156 (2006). doi: 10.1074/mcp.M500230-MCP200
9. J.C. Silva, R. Denny, C. Dorschel, M.V. Gorenstein, G.-Z. Li, K. Richardson, D. Wall and S.J. Geromanos, "Simultaneous qualitative and quantitative analysis of the *Escherichia coli* proteome: A sweet tale", *Mol. Cell. Proteomics* **5**, 589–607 (2006). doi: 10.1074/mcp.M500321-MCP200
10. T. Paalme, A. Kahru, R. Elken, K. Vanatalu, K. Tiisma and R. Vilu, "The computer-controlled continuous culture of *Escherichia coli* with smooth change of dilution rate (A-stat)", *J. Microbiol. Meth.* **24,** 145–153 (1995). doi: 10.1016/0167-7012(95)00064-X
11. U.K. Laemmli, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4", *Nature* **227,** 680–685 (1970). doi: 10.1038/227680a0
12. D.N. Perkins, D.J.C. Pappin, D.M. Creasy and J.S. Cottrell, "Probability-based protein identification by searching sequence databases using mass spectrometry data", *Electrophoresis* **20,** 3551–3567 (1999). doi: 10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2
13. P.R. Cutillas, B. Geering, M.D. Waterfield and B. Vanhaesebroeck, "Quantification of gel-separated proteins and their phosphorylation sites by LC-MS using unlabeled internal standards: Analysis of phosphoprotein dynamics in a B cell lymphoma cell line", *Mol Cell Proteomics* **4,** 1038–1051 (2005). doi: 10.1074/mcp.M500078-MCP200
14. V. Hatzimanikatis and K.H. Lee, "Dynamical analysis of gene networks requires both mRNA and protein expression information", *Metab. Eng.* **281,** 275–281 (1999). doi: 10.1006/mben.1999.0115
15. K.G. Le Roch, J.R. Johnson, L. Florens, Y. Zhou, A. Santrosyan, M. Grainger, S.F. Yan, K.C. Williamson, A.A. Holder, D.J. Carucci, J.R. Yates, III and E.A. Winzeler, "Global analysis of transcript and protein levels across the *Plasmodium falciparum* life cycle", *Genome Res.* **14,** 2308–2318 (2004). doi: 10.1101/gr.2523904
16. T.J. Griffin, S.P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood and R. Aebersold, "Complementary profiling of gene expression at the transcriptome and proteome levels in *Saccharomyces cerevisiae*", *Mol. Cell. Proteomics* **1,** 323–333 (2002). doi: 10.1074/mcp.M200001-MCP200
17. M.P. Washburn, A. Koller, G. Oshiro, R.R. Ulaszek, D. Plouffe, C. Deciu, E. Winzeler and J.R. Yates, III, "Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*", *Proc. Nat. Acad. Sci. USA* **100,** 3107–3112 (2003). doi: 10.1073/pnas.0634629100
18. Q. Tian, S.B. Stepaniants, M. Mao, L. Weng, M.C. Feetham, M.J. Doyle, E.C. Yi, H. Dai, V. Thorsson, J. Eng, D. Goodlett, J.P. Berger, B. Gunter, P.S. Linseley, R.B. Stoughton, R. Aebersold, S.J. Collins, W.A. Hanlon and L.E. Hood, "Integrated genomic and proteomic analyses of gene expression in mammalian cells", *Mol. Cell Proteomics* **3,** 960–969 (2004). doi: 10.1074/mcp.M400055-MCP200
19. R.W. Corbin, O. Paliy, F. Yang, J. Shabanowitz, M. Platt, C.E. Lyons, Jr, K. Root, J. McAuliffe, M.I. Jordan, S. Kustu, E. Soupene and D.F. Hunt. "Toward a protein profile of *Escherichia coli*: Comparison to its transcription profile", *Proc. Nat. Acad. Sci. USA* **100,** 9232–9237 (2003). doi: 10.1073/pnas.1533294100

20. R. Nahku, K. Valgepea, P. Lahtvee, S. Erm, K. Abner, K. Adamberg and R. Vilu, "Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures", *J. Biotechnol.* **145,** 60–65 (2010). doi: 10.1016/j.jbiotec.2009.10.007

PUBLICATION II

Valgepea K, Adamberg K, Nahku R, Lahtvee PJ, Arike L, Vilu R

Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase.

BMC Systems Biology, 4:166 (2010)

Valgepea et al. BMC Systems Biology 2010, 4:166
http://www.biomedcentral.com/1752-0509/4/166

BMC Systems Biology

RESEARCH ARTICLE Open Access

Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase

Kaspar Valgepea[1,2], Kaarel Adamberg[2,3], Ranno Nahku[1,2], Petri-Jaan Lahtvee[1,2], Liisa Arike[2,3], Raivo Vilu[1,2*]

Abstract

Background: The biotechnology industry has extensively exploited *Escherichia coli* for producing recombinant proteins, biofuels etc. However, high growth rate aerobic *E. coli* cultivations are accompanied by acetate excretion *i.e.* overflow metabolism which is harmful as it inhibits growth, diverts valuable carbon from biomass formation and is detrimental for target product synthesis. Although overflow metabolism has been studied for decades, its regulation mechanisms still remain unclear.

Results: In the current work, growth rate dependent acetate overflow metabolism of *E. coli* was continuously monitored using advanced continuous cultivation methods (A-stat and D-stat). The first step in acetate overflow switch (at $\mu = 0.27 \pm 0.02$ h^{-1}) is the repression of acetyl-CoA synthethase (Acs) activity triggered by carbon catabolite repression resulting in decreased assimilation of acetate produced by phosphotransacetylase (Pta), and disruption of the PTA-ACS node. This was indicated by acetate synthesis pathways PTA-ACKA and POXB component expression down-regulation before the overflow switch at $\mu = 0.27 \pm 0.02$ h^{-1} with concurrent 5-fold stronger repression of acetate-consuming Acs. This in turn suggests insufficient Acs activity for consuming all the acetate produced by Pta, leading to disruption of the acetate cycling process in PTA-ACS node where constant acetyl phosphate or acetate regeneration is essential for *E. coli* chemotaxis, proteolysis, pathogenesis etc. regulation. In addition, two-substrate A-stat and D-stat experiments showed that acetate consumption capability of *E. coli* decreased drastically, just as Acs expression, before the start of overflow metabolism. The second step in overflow switch is the sharp decline in cAMP production at $\mu = 0.45$ h^{-1} leading to total Acs inhibition and fast accumulation of acetate.

Conclusion: This study is an example of how a systems biology approach allowed to propose a new regulation mechanism for overflow metabolism in *E. coli* shown by proteomic, transcriptomic and metabolomic levels coupled to two-phase acetate accumulation: acetate overflow metabolism in *E. coli* is triggered by Acs down-regulation resulting in decreased assimilation of acetic acid produced by Pta, and disruption of the PTA-ACS node.

Background

Escherichia coli has not only been the prime organism for developing new molecular biology methods but also for producing recombinant proteins, low molecular weight compounds etc. in industrial biotechnology for decades due to its low cost manufacturing and end-product purification and its ability to reach high cell densities grown aerobically [1,2]. However, a major problem exists with aerobic *E. coli* cultivation on glucose at high growth rates-formation and accumulation of considerable amounts of acetic acid *i.e.* overflow metabolism. In addition to being detrimental for target product synthesis, accumulated acetate inhibits growth and diverts valuable carbon from biomass formation [3,4].

The acetate synthesis and utilization pathways [5] can be seen in Figure 1: acetate can be synthesized by

* Correspondence: raivo@kbfi.ee
[1]Tallinn University of Technology, Department of Chemistry, Akadeemia tee 15, 12618 Tallinn, Estonia
Full list of author information is available at the end of the article

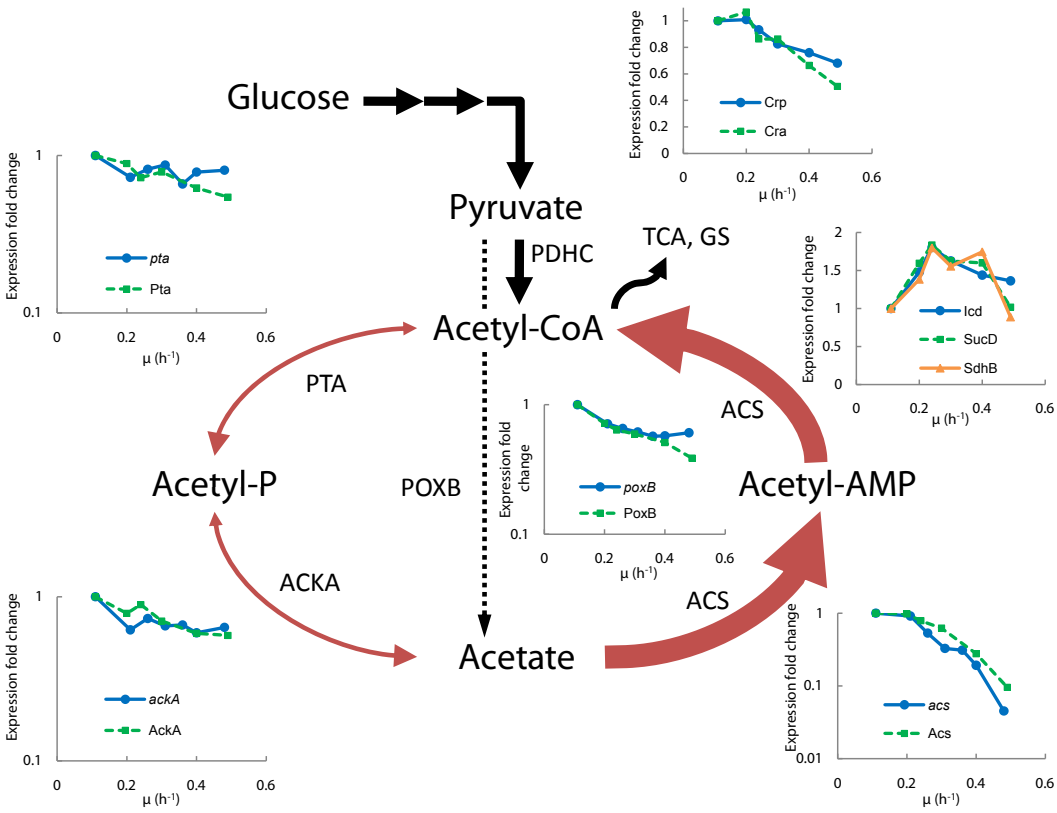


Figure 1 Effect of specific growth rate on acetate synthesis and utilization pathways, selected TCA cycle and carbon catabolite repressed gene and protein expression levels in *E. coli* A-stat experiments. PTA, phosphotransacetylase; ACKA, acetate kinase; ACS, acetyl-CoA synthetase; POXB, pyruvate oxidase; PDHC, pyruvate dehydrogenase complex; TCA, tricarboxylic acid cycle; GS, glyoxylate shunt; Crp, cyclic AMP receptor protein; Cra, catabolite repressor activator; Icd, isocitrate dehydrogenase; SucD, succinyl-CoA synthethase; SdhB, succinate dehydrogenase; μ, specific growth rate (h^{-1}). Thickness of red arrows denotes level of ACS and PTA-ACKA pathway repression (thick line represents stronger repression). Protein data points are average of two independent experiments, error bars are not shown for better visualization. Gene names are italicized. Refer to Additional file 2 for standard deviations and all the data.

phosphotransacetylase (PTA)/acetate kinase (ACKA) and by pyruvate oxidase (POXB). Acetic acid can be metabolized to acetyl-CoA either by the PTA-ACKA pathway or by acetyl-CoA synthetase (ACS) through an intermediate acetyl-AMP. The high affinity (K_m of 200 μM for acetic acid) ACS scavenges acetate at low concentrations whereas the low affinity PTA-ACKA pathway (K_m of 7-10 mM) is activated in the presence of high acetate concentrations [6].

The phenomenon of overflow metabolism has been studied widely over the years and it is commonly believed to be caused by an imbalance between the fluxes of glucose uptake and those for energy production and biosynthesis [7,8]. Several explanations such as the saturation of catalytic activities in the tricarboxylic acid (TCA) cycle [9,10] and respiratory chain [7,11,12], energy generation [5,13] or the necessity for coenzyme A replenishment [14] have been proposed. In addition to bioprocess level approaches [1,15], various genetic modifications of the acetate synthesis pathways extensively reviewed in De Mey *et al.* [15] have been made to minimize acetic acid production. For instance, it has been shown that deleting the main acetate synthesis route PTA-ACKA results in a strong reduction (up to 80%) of acetate excretion, maximum growth rate (*ca* 20%) and elevated levels of formate and lactate (*ca* 30-fold) [4,16-18], whereas *poxB* disruption causes reduction in biomass yield (*ca* 25%) and loss of aerobic growth efficiency of *E. coli* [19]. The latter indicates that acetate excretion cannot be simply excluded by

disrupting its synthesis routes without encountering other unwanted effects. Unfortunately, no clear conclusions could be drawn from batch experiments with an *acs* knock-out strain [4]. It should be noted that studies with *E. coli* genetically modified strains engineered to diminish acetate production in batch cultures have not fully succeeded in avoiding acetate accumulation together with increasing target product production yields and rates [15]. Additionally, these studies have not allowed elucidating the mechanism of overflow metabolism unequivocally [4,20,21].

Acetate overflow is a growth rate dependent phenomenon, but no study has specifically focused on growth rate dependency of protein and gene expression regulation, intra-and extracellular metabolite levels using also metabolic modeling. Describing the physiology of an organism on several 'omic levels is the basis of systems biology that facilitates better understanding of metabolic regulation [22]. In this study, *E. coli* metabolism at proteomic, transcriptomic and metabolomic levels was investigated using continuous cultivation methods prior to and after overflow metabolism was switched on. Usually, chemostat cultures are used for steady state metabolism analysis, however, we applied two changestat cultivation techniques: accelerostat (A-stat) and dilution rate stat (D-stat), see Methods section for details [23,24]. These cultivation methods were used as they provide three advantages over chemostat. Firstly, these changestat cultivation techniques precisely detect metabolically relevant switch points (*e.g.* start of overflow metabolism, maximum specific growth rate) and enable to monitor the dynamic patterns of several metabolic physiological responses simultaneously which could be left unnoticed using chemostat. Secondly, it is possible to collect vast amount of steady state comparable samples and by doing so, save time. Thirdly, both A-stat and D-stat enable to quantitatively study specific growth rate dependent co-utilization of growth substrates. Latter advantage was applied for investigating acetic acid consumption capability of *E. coli* at various dilution rates in this study. Combining changestat cultivation methods enables to study metabolism responses of the same genotype at different physiological states in detail without encountering the possible metabolic artifacts accompanied when using genetically modified strains.

Results obtained by studying specific growth rate dependent changes in *E. coli* proteome, transcriptome and metabolome in continuous cultures together with metabolic modeling allowed us to propose a new theory for acetate overflow: acetate excretion in *E. coli* is triggered by carbon catabolite repression mediated down-regulation of Acs resulting in decreased assimilation of acetate produced by Pta, and disruption of the PTA-ACS node.

Results

E. coli metabolic switch points characterization

In all accelerostat (A-stat) cultivation experiments, after the culture had been stabilized in chemostat at 0.10 h^{-1} to achieve steady state conditions, continuous increase in dilution rate with acceleration rate (a) 0.01 h^{-2} (0.01 h^{-1} per hour) was started. Continuous change of specific growth rate resulted in detecting several important changes in *E. coli* metabolism as demonstrated in Figure 2. Firstly, in A-stat cultivations where glucose was the only carbon source in the medium, acetic acid started to accumulate (*i.e.* overflow metabolism switch) at μ = 0.27 ± 0.02 h^{-1} (average ± standard deviation) and a two-phase acetate accumulation pattern was observed (discussed below; Figure 2). Cells reached maximum CO_2 production and O_2 consumption at μ = 0.46 ± 0.02 h^{-1} and metabolic fluctuations were observed at μ = 0.49 ± 0.03 h^{-1} followed by washout of culture at μ = 0.54 ± 0.03 h^{-1} (corresponding to maximum specific growth rate at given conditions). The nature of these fluctuations will be studied further and not covered in the current publication. All A-stat results were reproduced with relative standard deviation less than 10% with the exception of acetate production per biomass (Y_{OAc-}) (Table 1 and Figure S1 in Additional file 1).

Metabolomic responses to rising specific growth rate

A-stat cultivation enabled to study acetic acid accumulation profile in detail with increasing specific growth rate. Interestingly, a two-phase acetate accumulation pattern was observed (Figure 2). Slow accumulation of acetic

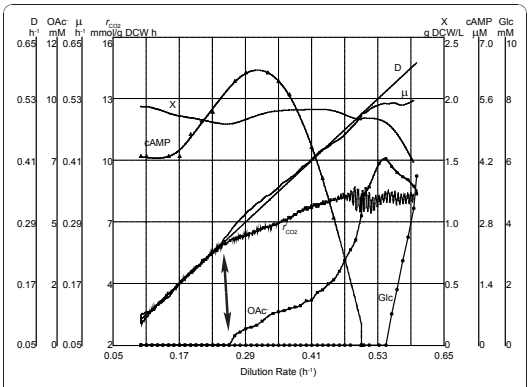


Figure 2 Increasing dilution rate dependent *E. coli* metabolism characterization in one A-stat cultivation (a = 0.01 h^{-2}). D, dilution rate (h^{-1}); X, biomass concentration (g dry cellular weight (DCW)/L); μ, specific growth rate (h^{-1}); r_{CO2}, specific CO_2 production rate (mmol/g DCW h); OAc^-, acetate concentration (mM); Glc, glucose concentration (mM); cAMP, cyclic AMP concentration (μM). Arrow indicates the start of overflow metabolism. Start of vertical axes was chosen for better visualization.

Table 1 A-stat and chemostat growth characteristics comparison and A-stat reproducibility over the studied specific growth rate range for three independent experiments

	μ = 0.24 h^{-1}		μ = 0.30 h^{-1}		μ = 0.40 h^{-1}		μ = 0.51 h^{-1}		μ = 0.10-0.47 h^{-1}
	Chemostat	A-stat	Chemostat	A-stat	Chemostat	A-stat	Chemostat	A-stat	A-stat RSD, %
Y_{XS}[a]	0.44	0.40 ± 0.01	0.46	0.41 ± 0.01	0.44	0.42 ± 0.00	0.43	0.41 ± 0.01	2.0
Y_{OAc-}[b]	NDE	NDE	0.53	0.90 ± 0.32	1.70	1.56 ± 0.23	3.25	3.35 ± 0.82	ND
Y_{cAMP}[c]	3.47	3.59 ± 0.39	3.25	3.55 ± 0.32	2.70	2.17 ± 0.07	0.86	0.71[e]	9.1
Y_{CO2}[d]	27.56	30.12 ± 2.04	27.55	27.19 ± 1.22	26.24	23.86 ± 1.41	ND	21.19 ± 0.19	5.6

A-stat values represent the average from three independent experiments and standard deviation follows the ± sign. Chemostat values from one experiment. NDE, not detected. ND, not determined. RSD, relative standard deviation.

[a]Biomass yield is given in g dry cell weight (DCW)/g glucose consumed (g DCW/g glucose).

[b]Acetic acid production per biomass is given in mmol acetic acid/g DCW.

[c]cAMP production per biomass is given in μmol cAMP/g DCW.

[d]Carbon dioxide (CO_2) production per biomass is given in mmol CO_2/g DCW.

[e]Data from one A-stat experiment.

acid started at μ = 0.27 ± 0.02 h^{-1} with concomitant change in specific CO_2 production rate (Figure 2). Faster accumulation of acetate was witnessed after cells had reached maximum CO_2 production at μ = 0.46 ± 0.02 h^{-1}. Quite surprisingly, production of the important carbon catabolite repression (CCR) signal molecule cAMP (Y_{cAMP}) rose from steady state chemostat level 2.45 ± 0.26 μmol/g dry cellular weight (DCW) (μ = 0.10 h^{-1}) to 3.55 ± 0.32 μmol/g DCW (μ = 0.30 h^{-1}) after which it sharply decreased to 1.30 ± 0.44 μmol/g DCW at μ = 0.45 h^{-1} (Figure S1 in Additional file 1). This abrupt decline took place simultaneously with the faster acetate accumulation profile described above (Figure 2 and Figure S1 in Additional file 1). In addition, similar two-phase acetate accumulation phenomenon was observed in a two-substrate (glucose + acetic acid) A-stat during the decrease of cAMP around specific growth rate 0.39 h^{-1} (Figure S2 in Additional file 1).

Significant fall in two of the measured pentose phosphate pathway intermediates ribose-5-phosphate (R5P) and erythrose-4-phosphate (E4P) was detected with increasing specific growth rate which could point to possible limitation in RNA biosynthesis during growth (Figure 3A). PTA-ACS node related compound nonesterified acetyl-CoA (HS-CoA) level declined two-fold simultaneously with cAMP after acetate started to accumulate (Figure 3B). This indicates the possible increase of other CoA containing compounds *e.g.* succinyl-CoA. Accumulation of TCA cycle intermediates α-ketoglutarate and isocitrate (Figure 3B) with increasing dilution rate could be associated with pyrimidine deficiency and decrease of ATP expenditure in the PTA-ACS cycle. Concurrently, intracellular concentrations of fructose-1,6-bisphosphate (FBP) and glyceraldehyde-3-phosphate (GAP) from the upper part of energy generating glycolysis increased 6- and 3-fold, respectively (Figure 3C).

Functional-genomic responses to rising specific growth rate

The two main known pathways for acetate synthesis phosphotransacetylase-acetate kinase (PTA-ACKA) and pyruvate oxidase (POXB) were down-regulated, both on gene and protein expression levels, from μ = 0.20 h^{-1} *i.e.* before acetate overflow was switched on. At the same time, there was a concurrent 10-fold repression of the acetic acid utilization enzyme acetyl-CoA synthetase (Acs). This substantial difference (5-fold) between the acetate synthesis and assimilation pathways expression suggests that the synthesized acetic acid cannot be fully assimilated with increasing growth rates (Figure 1).

We observed the beginning of carbon catabolite repression (CCR) induction prior to acetate accumulation in parallel with Acs down-regulation. This was indicated by down-regulation (3-fold on average) of CCR-mediated components: alternative (to glucose) substrate transport and utilization systems like galactose (MglAB), maltose (MalBEFKM), galactitol (GatABC), L-arabinose (AraF), D-ribose (RbsAB), C_4-dicarboxylates (DctA) and acetate (ActP, YjcH) (Figure 4C and Additional file 2). Moreover, expression of transcription activator Crp (cyclic AMP receptor protein which regulates the expression of Acs transcribing *acs-yjcH-actP* operon) and Cra (catabolite repressor activator; a global transcriptional protein essential for acetic acid uptake [25]) were reduced 1.5 and 2 times, respectively, in like manner to carbon catabolite repressed proteins mentioned above (Figure 1). Simultaneously, components of the gluconeogenesis pathway (Pck, MaeB, Pps) and glyoxylate shunt enzymes AceA, AceB (vital for acetate consumption) were repressed with growth rate increase (Figure 4B and Additional file 2). It should be emphasized that most of the TCA cycle gene and protein levels were maintained or even increased up to μ = 0.40 h^{-1}

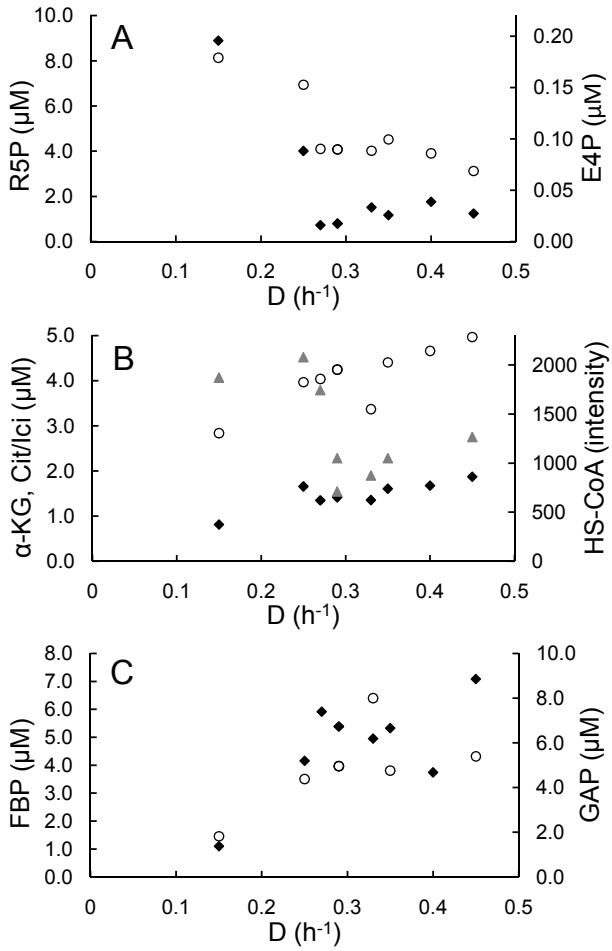


Figure 3 Dilution rate dependent intracellular metabolite patterns in one *E. coli* A-stat experiment. D, dilution rate (h^{-1}). (A) Pentose phosphate pathway metabolites. R5P, ribose-5-phosphate concentration (black diamond); E4P, erythrose-4-phosphate concentration (open circle). (B) TCA cycle metabolites and co-factor free CoA. α-KG, α-ketoglutarate concentration (black diamond); Cit/Ici, citrate/isocitrate pool concentration (open circle); HS-CoA, co-factor free CoA level (grey triangle). (C) Glycolysis (upper part) metabolites. FBP, fructose-1,6-bisphosphate concentration (black diamond); GAP, glyceraldehyde-3-phosphate concentration (open circle).

followed by sudden repression simultaneous to achieving maximum specific CO_2 production rate (μ = 0.46 ± 0.02 h^{-1}, see above; Figure 1 Figure 2 and Figure 4A). This may allude to no limitation at the TCA cycle level around the specific growth rate where overflow metabolism was switched on.

Acetic acid consumption capability studied by dilution rate stat (D-stat) and two-substrate A-stat cultivations

The beginning of a strong decrease in acetate assimilation enzyme Acs expression before overflow switch point implies to a possible connection between acetate assimilation capability and overflow metabolism of acetate (Figure 1). Therefore, specific growth rate dependent acetic acid consumption capabilities were investigated using D-stat and two-substrate A-stat methods. It was shown by D-stat experiments at various dilution rates that more than a 12-fold reduction in acetate consumption capability took place around overflow switch point, and its total loss was detected between dilution rates 0.45 and 0.505 ± 0.005 h^{-1} (Figure 5). Acetic acid consumption and production was also studied in a single experiment using two substrate (glucose + acetic acid) A-stat cultivation (Figure S2 in Additional file 1) which

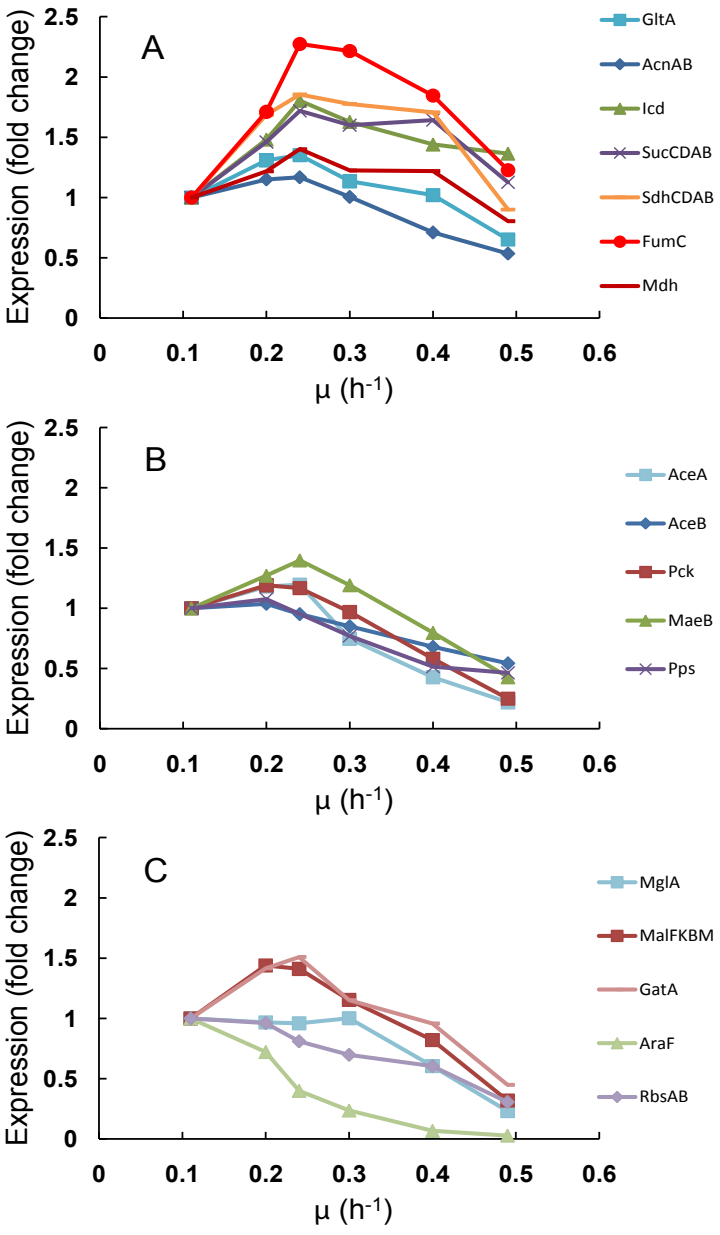


Figure 4 Specific growth rate dependent TCA cycle, glyoxylate shunt, glyconeogenesis and carbon catabolite repressed protein expression changes in *E coli* A-stat cultures. μ, specific growth rate (h^{-1}). (A) TCA cycle (average of proteins from the same operon are depicted as one point *e.g.* AcnAB). (B) Glyoxylate shunt (AceA, AceB) and glyconeogenesis. (C) Carbon catabolite repressed proteins. Protein data points are average of two independent experiments, error bars are not shown for better visualization (refer to Additional file 2 for standard deviations).

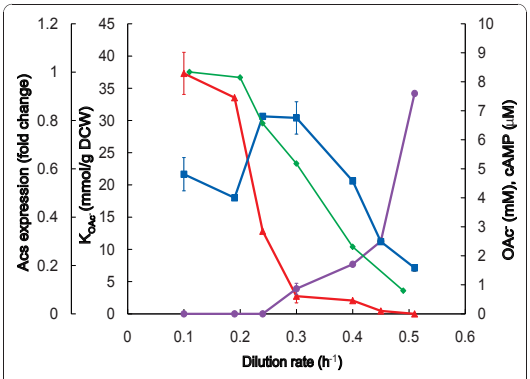


Figure 5 Dilution rate dependent acetic acid consumption capability in *E. coli* D-stat cultures. K_{OAc^-}, acetic acid consumption per biomass (red triangle); OAc^-, acetate concentration in chemostat before the start of acetic acid supplemented medium addition (violet circle); cAMP, cyclic AMP concentration (blue square); Acs, acetyl-CoA synthetase protein dilution rate dependent expression levels from A-stat (green diamond). Error bars represent the standard deviation from two independent D-stat experiments.

demonstrated that acetic acid consumption started to decrease at $\mu = 0.25\ h^{-1}$ and was completely abolished at $\mu = 0.48\ h^{-1}$ which fits into the range of dilution rates observed in D-stat.

A-stat comparison with chemostat

As could be seen from Table 1 major growth characteristics such as biomass yield (Y_{XS}), acetate (Y_{OAc^-}), cyclic AMP (Y_{cAMP}) and carbon dioxide (Y_{CO2}) production per biomass from A-stat and chemostat are all fully quantitatively comparable. The latter results enable to use A-stat data for quantitative modeling calculations. In addition, the two continuous cultivation methods were examined at transcriptome level using DNA microarrays. Transcript spot intensities from quasi steady state A-stat sample at $\mu = 0.48\ h^{-1}$ and chemostat sample at $\mu = 0.51\ h^{-1}$ showed an excellent Pearson product-moment correlation coefficient R = 0.964 (Figure S3 in Additional file 1; Additional file 3). This indicates good biological correlation between *E. coli* transcript profiles at similar specific growth rates in chemostat and A-stat. These results showed that our quasi steady state data from A-stat and D-stat cultures are steady state representative.

Proteome and transcriptome comparison

E. coli protein expression ratios for around 1600 proteins were generated by comparing two biological replicates at specific growth rates 0.20 ± 0.01; 0.26; 0.30 ± 0.01; 0.40 ± 0.00; 0.49 ± 0.01 h^{-1} with sample at $\mu = 0.10 \pm 0.01\ h^{-1}$ (chemostat point prior to the start of acceleration in A-stat) which produced Pearson correlation coefficients for two biological replicates in the indicated pairs of comparison in the range of R = 0.788-0.917 (Figure S4 in Additional file 1).

DNA microarray analysis of 4,321 transcripts was conducted with the Agilent platform using the samples from one A-stat cultivation. Gene expression ratios between specific growth rates 0.21; 0.26; 0.31; 0.36; 0.40; 0.48 h^{-1} and $\mu = 0.11\ h^{-1}$ (chemostat point prior to the start of acceleration in A-stat) were calculated. Comparison of gene and protein expression changes (between respective specific growth rates) revealed that components of the PTA-ACS node were regulated at transcriptional level as the absolute majority of the studied transcripts and proteins indicated by the good correlation between transcriptome and proteome expression profiles (Figure 1 and Figure S5 in Additional file 1).

Most recent studies have either failed to find a significant correlation between protein and mRNA abundances or have observed only a weak correlation (reviewed in [22]). It has been suggested that the main reasons for uncoupling of mRNA and protein abundances are protein regulation by post-translational modification, post-transcriptional regulation of protein synthesis, differences in the half-lives of mRNA and proteins, or possible functional requirement for protein binding [22]. As the cells in these studies were mostly cultured in non steady state condition, our steady state data with very good correlation between transcriptome and proteome implies that the physiological state of the culture (steady state vs. non steady state) could be an important factor in terms of mRNA and protein correlation determination. Transcriptome and proteome data are presented in Additional file 2 and at NCBI Gene Expression Omnibus and PRIDE database (see Methods for details), respectively.

Discussion

To gain better insights into the regulation of acetate overflow metabolism in *E. coli*, we studied specific growth rate dependent proteomic, transcriptomic and metabolomic patterns combined with metabolic modeling using advanced continuous cultivation methods, which has not been carried out before. Continuous monitoring of the specific growth rate effect on *E. coli* metabolism enabled us to precisely detect important metabolic shift points, the most important being the start of acetate overflow at $\mu = 0.27 \pm 0.02\ h^{-1}$ (Figure 2), and changing patterns of a number of important metabolites *e.g.* acetate, cAMP. Quite surprising was the down-regulation of the known acetate synthesis pathways, PTA-ACKA and POXB expression before overflow switch with increasing growth rate (Figure 1). A similar pattern has been seen before in chemostat cultures

but without emphasizing the possible physiological consequences [26-28]. A 10-fold repression of the acetic acid utilization enzyme acetyl-CoA synthetase (Acs) expression was observed concurrently with the down-regulation of the PTA-ACKA pathway indicating that acetic acid synthesis might exceed its assimilation (Figure 1). Our two substrate A-stat and D-stat experiments directly proved that acetate consumption capability of *E. coli* is specific growth rate dependent as acetate consumption started to decrease at $\mu = 0.25\ h^{-1}$ (Figure S2 in Additional file 1) and acetate consumption capability decreased 12-fold around overflow switch growth rate $\mu = 0.27 \pm 0.02\ h^{-1}$, respectively (Figure 5). In addition, it was shown that activation of carbon catabolite repression (CCR) and repression of Acs take place simultaneously prior to the start of overflow metabolism (Figure 1 Figure 4 and Figure 5). As a result, it is proposed that acetate overflow metabolism in *E. coli* is triggered by Acs down-regulation resulting in decreased assimilation of acetic acid produced by Pta, and disruption of the PTA-ACS node.

We showed that Acs was concurrently down-regulated five times more compared to the acetate synthesis pathways (Figure 1). In addition, the TCA cycle flux decrease as shown by change in CO_2 production at overflow switch growth rate indicates that carbon is not metabolized by the TCA cycle after the start of acetate accumulation with pre overflow switch rates (Figure 2 and Additional file 4). The latter is caused because the amount of acetyl-CoA entering the TCA cycle decreases after carbon is lost into excreted acetate. Stronger repression of the acetate consuming Acs in comparison with acetate synthesizing PTA-ACKA together with a decline in TCA cycle flux suggest disruption of acetic acid cycling at the PTA-ACS node (Figure 1). While this node may seem as a futile cycle, the fact is that numerous metabolic tasks involving the intermediate molecules of this cycle-acetyl phosphate (acetyl-P) and acetyl-AMP-are essential for proper *E. coli* growth (Figure 6). For instance, these molecules play a crucial role in bacterial chemotaxis regulation in which flagellar rotation is controlled by the activation level of the response regulator CheY [29] through either phosphotransfer from CheA [30,31] or acetyl-P [31,32], acetylation by acetyl-AMP [33,34] or co-regulation of both mechanisms [29]. It has been also demonstrated that acetyl-P synthesis is vital for EnvZ-independent regulation of outer membrane porins [35], pathogenesis [36] and regulation of several virulence factors [5]. Furthermore, it has been presented that acetyl-P interacts with phosphate concentration regulators PhoB-PhoR [37] and NRI protein which is part of a complex nitrogen sensing system [38]. Acetyl-P is critical for efficient degradation of unfolded or damaged proteins by ATP-dependent

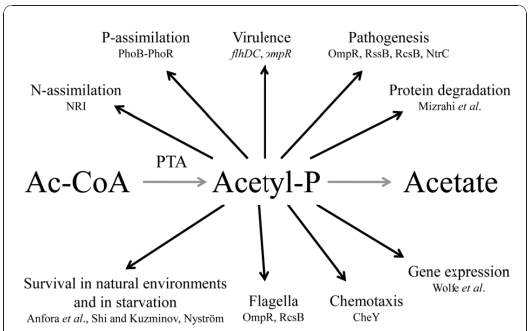


Figure 6 Acetyl-P as an important signal molecule. Ac-CoA, acetyl-CoA; PTA, phosphotransacetylase; refer to text for other abbreviations.

proteases [39]. Altogether, acetyl-P can influence the regulation of almost 100 other genes [40]. Finally, *pta* and/or *ackA* mutations were shown to affect repair-deficient *E. coli* mutants [41] and a *pta* mutant has been reported to be impaired in its ability to survive during glucose starvation, while the *ackA* mutant survived as well as the parent strain [42]. It is important to note that the only known pathway in *E. coli* for acetyl-P synthesis is the PTA-ACKA [5,31]. Taking all the previous into account, we conclude that acetyl-P as well as acetyl-AMP are essential for cellular growth of *E. coli*, and as acetic acid formation is the result of their dephosphorylation, acetic acid should be synthesized and consumed simultaneously during growth to maintain proper balance between metabolites of the PTA-ACS node. This is in agreement with Shin *et al.* [28] who proposed that wild-type *E. coli* constitutively synthesizes acetate even when growing on non-acetogenic carbon source succinate or at low growth rates in carbon limited cultures. It also has to be mentioned that acetic acid is a by-product in the synthesis of cysteine, methionine and arginine, covering around 0.4 mmol/g DCW (Additional file 4). Based on our experimental and literature data, production and re-assimilation of acetate might be over 1 mmol/g DCW at $\mu = 0.2\ h^{-1}$ (Text S2 in Additional file 1) which further supports the hypothesis of the necessity for constant acetic acid synthesis.

A similar regulation for overflow metabolism of acetate was posed for *Saccharomyces cerevisiae* by Postma and co-workers: they postulated that acetate accumulation is the result of insufficient acetyl-CoA synthetase activity for the complete functioning of the pyruvate dehydrogenase bypass because of glucose repression of ACS at high growth rates [43]. The hypothesis proposed here is also consistent with the observation that an *acs* mutant of *E. coli* accumulated acetate in chemostat cultures at dilution rate (D) $0.22\ h^{-1}$ whereas acetate

overflow was started in wild-type at a higher D = 0.35 h^{-1} [28]. Furthermore, it has been shown that over-expression of *acs* [44] and constitutively expressed *acs* together with glyoxylate shunt repressors *iclR* and *fadR* mutant resulted in a significant reduction in acetate accumulation in glucose batch fermentations [28]. Adams and co-workers showed that as a result of micro-evolution, *E. coli* increased acetate consumption capability by over-expressing Acs (not AckA) [45,46], further supporting the connection between Acs activity and acetate accumulation.

As Acs down-regulation is responsible for triggering overflow metabolism and the resulting accumulation of acetate is detrimental to cellular growth, it bears questioning why *E. coli* has not evolved towards maintaining sufficient Acs levels for acetate assimilation in all growth conditions. Growth conditions in *E. coli* native environments are rough as concentrations of utilizable carbon sources including acetate are in the low mg L^{-1} range and access to nutrients is troublesome [47]. These harsh conditions force *E. coli* to make its metabolism ready for scavenging all possible carbon sources including acetate. However, in nutrient rich laboratory conditions, *E. coli* focuses on anthropic growth [48] and biomass production rate, primarily realized by enhancing readily oxidizable substrate (glucose) uptake kinetics which in turn results in Acs repression through CCR and thus, acetate accumulation [46]. This indicates that an active Acs is not essential for rapid growth for *E. coli*. It seems that maintaining high expression levels of acetate assimilation components (and also other alternative substrates ones) is energetically not favorable at higher growth rates. Moreover, as the space on cell membrane is limited and as *E. coli* achieves more rapid growth probably by increasing the number of glucose transport machinery components on the membrane, using area for alternative substrate transport proteins is not beneficial for faster growth. Interestingly, even in one of its natural environments-urinary tract-where a continuous dilution of acetate occurs, it has been shown that metabolizing acetate to acetyl-CoA by Acs is not essential for normal *E. coli* colonization as PTA-ACKA pathway and maintenance of a proper intracellular acetyl-P concentration are necessary for colonizing murine urinary tract [32].

Based on all the points discussed above, PTA-ACS might function as a futile cycle to provide rapid regulation of acetyl-P concentration in the cell for an active chemotaxis that is vital in natural nutrient-depleted environments, fighting against other organisms (acetate production), pathogenesis, biofilm formation etc. This hypothesis is consistent with the fact that the flagellar assembly and regulation operon (*tar-tap-cheRBYZ*) was more intensively expressed at lower growth rates (Additional file 2) where residual glucose concentration is smaller.

Concerning Acs down-regulation, it is possible that CCR is responsible for its repression as proposed by Treves *et al.* [46] showing the link between ACS expression level and acetate accumulation. In our experiments, it was shown that activation of CCR and repression of Acs take place simultaneously prior to the start of overflow metabolism (Figure 1 and Figure 4). As it is well known that CCR is initiated by the presence of glucose in the medium [49,50], we propose that increasing residual glucose concentration accompanying smooth rise of dilution rate in A-stat triggers Acs down-regulation by CCR. The cAMP-Crp complex is one of the major players in CCR of *E. coli* as cAMP binding to Crp drastically increases its affinity towards activating the promotors of catabolic enzymes, including Acs [6,49,50]. We measured a 1.5-fold decrease in Crp expression with increasing growth rate (Figure 1) that is in agreement with the data in the literature [51]. In addition, when *E. coli* mutant defective in the gene *crp* was cultivated in glucose-limited chemostat at a low D = 0.10 h^{-1}, it accumulated acetate whereas the wild-type did not [52]. Furthermore, it exhibited a 34% higher biomass yield relative to the wild-type-this increase might be explained by reduced ATP wasting in the acetate futile cycle, which can be directed to biomass synthesis. Moreover, Khankal *et al.* [53] noted that *E. coli* CRP* mutants that do not require Crp binding to cAMP to activate the expression of catabolic genes showed lowered glucose effect on xylose consumption, 3.6 times higher *acs* expression levels and secreted substantially less acetate in xylitol producing batch fermentations. The connection between cAMP concentration and acetic acid consumption capability, together with the two-phase acetate accumulation profile observed in A-stat and D-stat cultures (Figure 2 and Figure 5) suggests a correlation between increasing residual glucose concentration mediated cAMP-Crp repression and acetate accumulation. Thus, cAMP-Crp dependent regulation of Acs transcribing *acs-yjcH-actP* operon might be a reason for acetate excretion, as also proposed by Veit *et al.* [10]. Our hypothesis of the CCR mediated acetate overflow metabolism is as well in agreement with the fact that rising glucose lowers the intracellular Crp level through the autoregulatory loop of the *crp* gene [54]. However, other mechanisms can also be involved in Acs down-regulation, for example by Cra (Figure 1). Indeed, Sarkar and colleagues have shown that glucose uptake and acetate production rates increased with a decrease of acetate consumption in an *E. coli cra* mutant [55].

What could be the biological relevance of the disruption of the PTA-ACS node? Firstly, decline of the ATP-spending PTA-ACS cycle throughput with increasing growth rate points to possible lower ATP spilling (our model calculations). Secondly, disruption of the PTA-

ACS node decreases the energy needed for expression of this cycle's components. As the disruption of PTA-ACS cycle is CCR-mediated, repression of other alternative substrate transport and utilization enzymes by CCR enables to save additional energy. This could all lead to the decrease of ATP production as was indicated by the diminishing TCA cycle fluxes (Figure 2). Hence, it is plausible that cells repress (by CCR) the expression levels of alternative substrate utilization components (including Acs) for making space on the cell membrane for more preferred substrate (glucose) utilization and ATP producing components to achieve faster growth (see above).

Finally, it was demonstrated that highly reproducible A-stat data are well comparable to chemostat at the level of major growth characteristics and transcriptome, hence quasi steady state data from A-stat can be considered steady state representative (Table 1; Figure S1 and Figure S3 in Additional file 1). Furthermore, as shown also by Postma *et al*. for *S. cerevisiae* [43], chemostat is not fully suitable for characterization of dilution rate dependent metabolic transitions, whereas A-stat should be considered an appropriate tool for this. A-stat is especially well suited for the studies of the details of transient metabolism processes. Dynamic behavior of acetate, cAMP etc. with increasing specific growth rate (Figure 2 Figure 3 and Figure S1 in Additional file 1) and change in acetic acid consumption capability in the two-substrate A-stat (Figure S2 in Additional file 1) could be cited as good examples of the latter.

Conclusion

This study is an excellent example of how a systems biology approach using highly reproducible advanced cultivation methods coupled with multiple 'omics analysis and metabolic modeling allowed to propose a new possible regulation mechanism for overflow metabolism in *E. coli*: acetate overflow is triggered by carbon catabolite repression mediated Acs down-regulation resulting in decreased assimilation of acetate produced by Pta, and disruption of the PTA-ACS node. The practical implications derived from this could lead to better engineering of *E. coli* in overcoming several metabolic obstacles, increasing production yields etc.

Methods

Bacterial strain, medium and continuous cultivation conditions

The *E. coli* K12 MG1655 (λ^- F^- *rph-1Fnr*$^+$; Deutsche Sammlung von Mikroorganismen und Zellkulturen, Germany) strain was used in all experiments. Growth and physiological characteristics in accelerostat (A-stat) cultivations were determined using a defined minimal medium as described before by Nahku *et al*. [51], except 4.5 g/L α-(D)-glucose and 100 μl L^{-1} Antifoam C (Sigma Aldrich, St. Louis, LO) was used. The latter was also used in dilution rate stat (D-stat) experiments as the main cultivation medium. In addition, a second medium was used in D-stat where the main medium was supplemented by acetic acid and prepared as follows: 300 ml medium was withdrawn from the main cultivation medium and supplemented with 3 ml of glacial acetic acid (99.9%). One A-stat experiment (referred to as two-substrate A-stat) was carried out with the same medium as other A-stats, but in addition supplemented with acetic acid (final concentration 5 mM).

The continuous (both A-stat and D-stat) cultivation system consisted of 1.25 L Biobundle bioreactor (Applikon Biotechnology B.V., Schiedam, the Netherlands) controlled by an ADI 1030 biocontroller (Applikon Biotechnology B.V.) and a cultivation control program "BioXpert NT" (Applikon Biotechnology B.V.). The system was equipped with OD, pH, pO_2, CO_2 and temperature sensors. The bioreactor was set on a balance whose output was used as the control variable to ensure constant culture volume (300 ± 1 mL). Similarly, the inflow was controlled through measuring the mass of the fresh culture medium.

A-stat cultivation system and control algorithms used are described in more detail in our previous works [24,51,56]. Dilution rate stat (D-stat) is a continuous cultivation method where dilution rate is constant as in a chemostat while an environmental parameter is smoothly changed [24]. The D-stat experiments in this study were carried out with a slight modification: instead of changing an environmental parameter, two different media were used to keep dilution rate constant. After achieving steady state conditions in chemostat using minimal medium supplemented with glucose, addition of the second medium complemented with glucose and acetic acid was started. The feeding rate of the initial medium was decreased at the same time, resulting in constant glucose concentration in the feed. The acetic acid concentration in the bioreactor as a result of inflow has to be determined to enable precise acetic acid consumption/production rate calculation for the bacteria. Hence, increase of acetic acid concentration in bioreactor was calculated and validated in duplicate non-inoculatedD-stat test experiments producing an average standard deviation of 1.24 mM between calculated and measured acetic acid concentrations.

All continuous cultivation experiments were carried out at 37°C, pH 7 and under aerobic conditions (air flow rate 150 ml min^{-1}) with an agitation speed of 800 rpm. Four A-stat cultivations were performed with acceleration rate (a) 0.01 h^{-2}. Duplicate D-stat experiments were performed at dilution rates 0.10; 0.30; 0.505 ± 0.005 h^{-1} and single experiments at 0.19; 0.24;

0.40; 0.45 h^{-1}. The acetic acid addition profile was set to achieve 32 ± 6 mM and 58 ± 5 mM in 7 hours inside the bioreactor for experiments at dilution rates 0.10-0.24 h^{-1} and 0.30-0.51 h^{-1}, respectively. The growth characteristics of the bacteria were calculated on the basis of total volume of medium pumped out from bioreactor (L), biomass (g DCW), organic acid concentrations in culture medium (mM) and CO_2 concentration in the outflow gas (mM). Formulas were as described in a previous study [24]. It should be noted that the absolute CO_2 concentrations could be error-prone due to measurement difficulties. However, this does not influence the dynamic pattern of specific CO_2 production rate (r_{CO2}) during specific growth rate increase.

Analytical methods

The concentrations of organic acids (lactate, acetate and formate), ethanol and glucose in the culture medium were determined by HPLC and cellular dry weight (expressed as DCW) as described by Nahku *et al.* [51].

Protein expression analysis

Refer to Text S1 in Additional file 1 for detailed description. Shortly, protein expression ratios for around 1600 proteins (identified for each growth rate at a > 95% confidence interval in average from 89,303 distinct 2 or more high-confidence peptides) were generated from mass spectrometric spectra by firstly calculating the ratios between continuous cultivation samples at specific growth rates 0.10 ± 0.01 h^{-1} (chemostat point prior to the start of acceleration in A-stat); 0.20 ± 0.01; 0.26; 0.30 ± 0.01; 0.40 ± 0.00; 0.49 ± 0.01 h^{-1} and batch sample grown on medium containing $^{15}NH_4Cl$ as the only source of ammonia. Secondly, the ratios between the mentioned specific growth rates with chemostat point (μ = 0.10 ± 0.01 h^{-1}) for two biological replicates were calculated to yield protein expression levels for respective specific growth rates. Protein (and gene) expression measurement results are shown in Additional file 2. Proteomic analysis data is also available at the PRIDE database [57]http://www.ebi.ac.uk/pride under accession numbers 12189-12199 (username: review74613, password: Ge9T48e8). The data was converted using PRIDE Converter http://code.google.com/p/pride-converter [58].

Gene expression profiling

DNA microarray analysis of 4,321 transcripts was conducted with the Agilent platform using the data from one A-stat cultivation (a = 0.01 h^{-2}), and gene expression ratios between specific growth rates 0.21; 0.26; 0.31; 0.36; 0.40; 0.48 h^{-1} and μ = 0.11 h^{-1} were calculated. Transcript spot intensities of chemostat sample (sample from D-stat prior to acetic acid addition) from μ = 0.51 h^{-1} and A-stat μ = 0.48 h^{-1} were used for the two method's comparison at transcriptome level. Gene (and protein) expression measurement results are shown in Additional file 2. DNA microarray data is also available at NCBI Gene Expression Omnibus (Reference series: GSE23920). The details of the procedure are provided in Text S1 in Additional file 1.

Metabolome analysis

Sampling was carried out by the rapid centrifugation method. Acquity UPLC (Waters, Milford, MA) together with end-capped HSS C18 T3 1.8 μm, 2.1 × 100 mm column for compound separation coupled to TOF-MS with an electrospray ionization (ESI) source was used for detection (LCT Premiere, Waters). The details of the procedure are provided in Text S1 in Additional file 1.

Additional material

Additional file 1: Detailed Methods (Text S1); calculation of acetate reconsumption (Text S2); Supplementary Figures S1-S5.

Additional file 2: Growth rate dependent gene (one A-stat) and average protein expression changes of two A-stat experiments with *Escherichia coli* K12 MG1655. Transcriptome and proteome analysis results, also with standard deviations.

Additional file 3: Gene spot intensities of A-stat at μ = 0.48 h^{-1} and chemostat at μ = 0.51 h^{-1} experiments with *Escherichia coli* K12 MG1655. Data for A-stat and chemostat transcriptome comparison.

Additional file 4: Simplified metabolic flux analysis. Detailed description of model calculations with simplified metabolic flux analysis.

Acknowledgements

The financial support for this research was provided by the Enterprise Estonia project EU29994, and Ministry of Education, Estonia, through the grant SF0140090s08. The authors would like to thank Lauri Peil and Elina Pelonen for help in carrying out 'omics analysis.

Author details

[1]Tallinn University of Technology, Department of Chemistry, Akadeemia tee 15, 12618 Tallinn, Estonia. [2]Competence Centre of Food and Fermentation Technologies, Akadeemia tee 15b, 12618 Tallinn, Estonia. [3]Tallinn University of Technology, Department of Food Processing, Ehitajate tee 5, 19086 Tallinn, Estonia.

Authors' contributions

KV, KA, and RV drafted the manuscript. RN, PL, and LA helped in preparing the manuscript. KV, RN, and PL designed and performed the experiments. KV analysed the experimental data. RN, PL, and LA carried out the 'omics analysis. KV, KA, and RV guided and coordinated the project. All authors read and approved the manuscript.

Received: 16 June 2010 Accepted: 1 December 2010
Published: 1 December 2010

References

1. Eiteman MA, Altman E: **Overcoming acetate in *Escherichia coli* recombinant protein fermentations.** *Trend Biotechnol* 2006, **24**:530-536.
2. Clomburg JM, Gonzalez R: **Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology.** *Appl Microbiol Biotechnol* 2010, **86**:419-434.

3. Nakano K, Rischke M, Sato S, Märkl H: **Influence of acetic acid on the growth of *Escherichia coli* K12 during high-cell-density cultivation in a dialysis reactor.** *Appl Microbiol Biotechnol* 1997, **48**:597-601.
4. Contiero J, Beatty CM, Kumari S, DeSanti CL, Strohl WR, Wolfe AJ: **Effects of mutations in acetate metabolism in high-cell-density growth of *Escherichia coli*.** *J Ind Microbiol Biotechnol* 2000, **24**:421-430.
5. Wolfe AJ: **The acetate switch.** *Microbiol Mol Biol Rev* 2005, **69**:12-50.
6. Kumari S, Beatty CM, Browning DF, Busby SJ, Simel EJ, Hovel-Miner G, Wolfe AJ: **Regulation of acetyl coenzyme A synthetase in *Escherichia coli*.** *J Bacteriol* 2000, **182**:4173-4179.
7. Han K, Lim HC, Hong J: **Acetic acid formation in *Escherichia coli* fermentation.** *Biotechnol Bioeng* 1992, **39**:663-671.
8. Farmer WR, Liao JC: **Reduction of aerobic acetate production by *Escherichia coli*.** *Appl Environ Microbiol* 1997, **63**:3205-3210.
9. Majewski RA, Domach MM: **Simple constrained-optimization view of acetate overflow in *E. coli*.** *Biotechnol Bioeng* 1990, **35**:732-738.
10. Veit A, Polen T, Wendisch V: **Global gene expression analysis of glucose overflow metabolism in *Escherichia coli* and reduction of aerobic acetate formation.** *Appl Microbiol Biotechnol* 2007, **74**:406-421.
11. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.** *Appl Environ Microbiol* 1994, **60**:3724-3731.
12. Paalme T, Elken R, Kahru A, Vanatalu K, Vilu R: **The growth rate control in *Escherichia coli* at near to maximum growth rates: the A-stat approach.** *Antonie van Leeuwenhoek* 1997, **71**:217-230.
13. Kayser A, Weber J, Hecht V, Rinas U: **Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state.** *Microbiology* 2005, **151**:693-706.
14. El-Mansi M: **Flux to acetate and lactate excretions in industrial fermentations: Physiological and biochemical implications.** *J Ind Microbiol Biotechnol* 2004, **31**:295-300.
15. De Mey M, De Maeseneire S, Soetaert W, Vandamme E: **Minimizing acetate formation in *E. coli* fermentations.** *J Ind Microbiol Biotechnol* 2007, **34**:689-700.
16. El-Mansi EM, Holms WH: **Control of carbon flux to acetate excretion during growth of *Escherichia coli* in batch and continuous cultures.** *J Gen Microbiol* 1989, **135**:2875-2883.
17. Yang Y-T, Bennett GN, San K-Y: **Effect of inactivation of *nuo* and *ackA-pta* on redistribution of metabolic fluxes in *Escherichia coli*.** *Biotechnol Bioeng* 1999, **65**:291-297.
18. Dittrich CR, Vadali RV, Bennett GN, San K-Y: **Redistribution of metabolic fluxes in the central aerobic metabolic pathway of *E. coli* mutant strains with deletion of the *ackA-pta* and *poxB* pathways for the synthesis of isoamyl acetate.** *Biotechnol Prog* 2005, **21**:627-631.
19. Abdel-Hamid AM, Attwood MM, Guest JR: **Pyruvate oxidase contributes to the aerobic growth efficiency of *Escherichia coli*.** *Microbiology* 2001, **147**:1483-1498.
20. Phue J, Noronha SB, Hattacharyya R, Wolfe AJ, Shiloach J: **Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and Northern blot analyses.** *Biotechnol Bioeng* 2005, **90**:805-820.
21. Castaño-Cerezo S, Pastor JM, Renilla S, Bernal V, Iborra JL, Cánovas M: **An insight into the role of phosphotransacetylase (*pta*) and the acetate/acetyl-CoA node in *Escherichia coli*.** *Microb Cell Fact* 2009, **8**:54.
22. Zhang W, Li F, Nie L: **Integrating multiple 'omics' analysis for microbial biology: application and methodologies.** *Microbiology* 2010, **156**:287-301.
23. Paalme T, Kahru A, Elken R, Vanatalu K, Tiisma K, Vilu R: **The computer-controlled continuous culture of *Escherichia coli* with smooth change of dilution rate.** *J Microbiol Methods* 1995, **24**:145-153.
24. Kasemets K, Drews M, Nisamedtinov I, Adamberg K, Paalme T: **Modification of A-stat for the characterization of microorganisms.** *J Microbiol Methods* 2003, **55**:187-200.
25. Saier MH, Ramseier TO: **The Catabolite Repressor/Activator (Cra) Protein of Enteric Bacteria.** *J Bacteriol* 1996, **178**:3411-3417.
26. Vemuri GN, Altman E, Sangurdekar DP, Khodursky AB, Eiteman MA: **Overflow metabolism in *Escherichia coli* during steady-state growth: transcriptional regulation and effect of the redox ratio.** *Appl Environ Microbiol* 2006, **72**:3653-3661.
27. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M: **Multiple high-throughput analyses monitor the response of *E. coli* to perturbations.** *Science* 2007, **316**:593-597.
28. Shin S, Chang D, Pan JG: **Acetate Consumption Activity Directly Determines the Level of Acetate Accumulation During *Escherichia coli* W3110 Growth.** *J Microbiol Biotechnol* 2009, **19**:1127-1134.
29. Barak R, Abouhamad WN, Eisenbach M: **Both acetate kinase and acetyl coenzyme A synthetase are involved in acetate-stimulated change in the direction of flagellar rotation in *Escherichia coli*.** *J Bacteriol* 1998, **180**:985-988.
30. Da Re SS, Deville-Bonne D, Tolstykh T, V ron M, Stock JB: **Kinetics of CheY phosphorylation by small molecule phosphodonors.** *FEBS Lett* 1999, **457**:323-326.
31. Mayover TL, Halkides CJ, Stewart RC: **Kinetic characterization of CheY phosphorylation reactions: comparison of P-CheA and small-molecule phosphodonors.** *Biochemistry* 1999, **38**:2259-2271.
32. Klein AH, Shulla A, Reimann Sa, Keating DH, Wolfe AJ: **The intracellular concentration of acetyl phosphate in *Escherichia coli* is sufficient for direct phosphorylation of two-component response regulators.** *J Bacteriol* 2007, **189**:5574-5581.
33. Barak R, Welch M, Yanovsky A, Oosawa K, Eisenbach M: **Acetyladenylate or its derivative acetylates the chemotaxis protein CheY *in vitro* and increases its activity at the flagellar switch.** *Biochemistry* 1992, **31**:10099-10107.
34. Yan J, Barak R, Liarzi O, Shainskaya A, Eisenbach M: ***In vivo* acetylation of CheY, a response regulator in chemotaxis of *Escherichia coli*.** *J Mol Biol* 2008, **376**:1260-1271.
35. Matsubara M, Mizuno T: **EnvZ-independent phosphotransfer signaling pathway of the OmpR-mediated osmoregulatory expression of OmpC and OmpF in *Escherichia coli*.** *Biosci Biotechnol Biochem* 1999, **63**:408-414.
36. Anfora AT, Halladin DK, Haugen BJ, Welch RA: **Uropathogenic *Escherichia coli* CFT073 is adapted to acetatogenic growth but does not require acetate during murine urinary tract infection.** *Infect Immun* 2008, **76**:5760-5767.
37. McCleary W, Stock J: **Acetyl phosphate and the activation of 2-component response regulators.** *J Biol Chem* 1994, **269**:31567-31572.
38. Feng J, Atkinson MR, McCleary W, Stock JB, Wanner BL, Ninfa AJ: **Role of phosphorylated metabolic intermediates in the regulation of glutamine synthetase synthesis in *Escherichia coli*.** *J Bacteriol* 1992, **174**:6061-6070.
39. Mizrahi I, Biran D, Ron EZ: **Involvement of the Pta-AckA pathway in protein folding and aggregation.** *Res Microbiol* 2009, **160**:80-84.
40. Wolfe AJ, Chang D-E, Walker JD, Seitz-Partridge JE, Vidaurri MD, Lange CF, Prüß BM, Henk MC, Larkin JC, Conway T: **Evidence that acetyl phosphate functions as a global signal during biofilm development.** *Mol Microbiol* 2003, **48**:977-988.
41. Shi IY, Kuzminov A: **A Defect in the Acetyl Coenzyme-Acetate Pathway Poisons Recombinational Repair-Deficient Mutants of *Escherichia coli*.** *J Bacteriol* 2005, **187**:1266-1275.
42. Nyström T: **The glucose-starvation stimulon of *Escherichia coli*: induced and repressed synthesis of enzymes of central metabolic pathways and role of acetyl phosphate in gene expression and starvation survival.** *Mol Microbiol* 1994, **12**:833-843.
43. Postma E, Verduyn C, Scheffers Wa, Van Dijken JP: **Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*.** *Appl Environ Microbiol* 1989, **55**:468-477.
44. Lin H, Castro NM, Bennett GN, San K: **Acetyl-CoA synthetase overexpression in *Escherichia coli* demonstrates more efficient acetate assimilation and lower acetate accumulation: a potential tool in metabolic engineering.** *Appl Microbiol Biotechnol* 2006, **71**:870-874.
45. Rosenzweig F, Adams J: **Microbial Evolution in a Simple Unstructured Environment: Genetic Differentiation in *Escherichia coli*.** *Genetics* 1994, **917**:903-917.
46. Treves DS, Manning S, Adams J: **Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*.** *Mol Biol Evol* 1998, **15**:789-797.
47. Franchini AG, Egli T: **Global gene expression in *Escherichia coli* K-12 during short-term and long-term adaptation to glucose-limited continuous culture conditions.** *Microbiology* 2006, **152**:2111-2127.

48. Hardiman T, Lemuth K, Keller M, Reuss M, Siemannherzberg M: **Topology of the global regulatory network of carbon limitation in *Escherichia coli*.** *J Biotechnol* 2007, **132**:359-374.
49. Görke B, Stülke JR: **Carbon catabolite repression in bacteria: many ways to make the most out of nutrients.** *Nat Rev Microbiol* 2008, **6**:613-624.
50. Narang A: **Quantitative effect and regulatory function of cyclic adenosine 5⊠-phosphate in *Escherichia coli*.** *J Biosci* 2009, **34**:445-463.
51. Nahku R, Valgepea K, Lahtvee PJ, Erm S, Abner K, Adamberg K, Vilu R: **Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures.** *J Biotechnol* 2010, **145**:60-65.
52. Nanchen A, Schicker A, Revelles O, Sauer U: **Cyclic AMP-dependent catabolite repression is the dominant control mechanism of metabolic fluxes under glucose limitation in *Escherichia coli*.** *J Bacteriol* 2008, **190**:2323-2330.
53. Khankal R, Chin JW, Ghosh D, Cirino PC: **Transcriptional effects of CRP* expression in *Escherichia coli*.** *J Biol Eng* 2009, **3**:13.
54. Ishizuka H, Hanamura A, Inada T, Aiba H: **Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the crp gene.** *EMBO J* 1994, **13**:3077-3082.
55. Sarkar D, Siddiquee KA, Araúzo-Bravo MJ, Oba T, Shimizu K: **Effect of *cra* gene knockout together with *edd* and *iclR* genes knockout on the metabolism in *Escherichia coli*.** *Arch Microbiol* 2008, **190**:559-751.
56. Adamberg K, Lahtvee PJ, Valgepea K, Abner K, Vilu R: **Quasi steady state growth of *Lactococcus lactis* in glucose-limited acceleration stat (A-stat) cultures.** *Antonie van Leeuwenhoek* 2009, **95**:219-226.
57. Martens L, Hermjakob H, Jones P, Adamski M, Taylor C, States D, Gevaert K, Vandekerckhove J, Apweiler R: **PRIDE: the proteomics identifications database.** *Proteomics* 2005, **5**:3537-3545.
58. Barsnes H, Vizcaíno JA, Eidhammer I, Martens L: **PRIDE Converter: making proteomics data-sharing easy.** *Nat Biotechnol* 2009, **27**:598-599.

doi:10.1186/1752-0509-4-166
Cite this article as: Valgepea *et al.*: Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Systems Biology* 2010 **4**:166.

PUBLICATION III

Arike L, Valgepea K, Peil L, Nahku R, Adamberg K, Vilu R

Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*

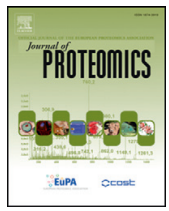
Journal of Proteomics, 75(17):5437-5448 (2012)

JOURNAL OF PROTEOMICS 75 (2012) 5437–5448

Available online at www.sciencedirect.com

SciVerse ScienceDirect

www.elsevier.com/locate/jprot



Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*

L. Arike[a,b,c,*], *K. Valgepea*[a,d], *L. Peil*[c,e], *R. Nahku*[a,d], *K. Adamberg*[a,b], *R. Vilu*[a,d]

[a]*Competence Center of Food and Fermentation Technologies, Akadeemia tee 15A, 12618 Tallinn, Estonia*
[b]*Tallinn University of Technology, Department of Food Processing, Ehitajate tee 5, 19086 Tallinn, Estonia*
[c]*University of Tartu, Institute of Technology, Proteomics Core Laboratory, Nooruse 1, 50411 Tartu, Estonia*
[d]*Tallinn University of Technology, Department of Chemistry, Akadeemia tee 15, 12618 Tallinn, Estonia*
[e]*Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, UK*

ARTICLE INFO

Article history:
Received 2 March 2012
Accepted 24 June 2012
Available online 5 July 2012

Keywords:
Label-free absolute
quantitative proteomics
Spectral counting
Precursor ion intensity
APEX
emPAI
iBAQ

ABSTRACT

Three different label-free proteome quantification methods – APEX, emPAI and iBAQ – were evaluated to measure proteome-wide protein concentrations in the cell. All the methods were applied to a sample from *Escherichia coli* chemostat culture. A Pearson squared correlation of approximately 0.6 among the three quantification methods was demonstrated. Importantly, the sum of quantified proteins by iBAQ and emPAI corresponded with the Lowry total protein quantification, demonstrating applicability of label-free methods for an accurate calculation of protein concentrations at the proteome level. The iBAQ method showed the best correlation between biological replicates, a normal distribution among all protein abundances, and the lowest variation among ribosomal protein abundances, which are expected to have equal amounts.

Absolute quantitative proteome data enabled us to evaluate metabolic cost for protein synthesis and apparent catalytic activities of enzymes by integration with flux analysis. All the methods demonstrated similar ATP costs for protein synthesis for different cellular processes and that costs for expressing biomass synthesis related proteins were higher than those for energy generation. Importantly, catalytic activities of energy metabolism enzymes were an order or two higher than those of monomer synthesis. Interestingly, a staircase-like protein expression was demonstrated for most of the transcription units.

1. Introduction

Quantitative proteomics has become a standard method in biological studies to measure cellular responses to environmental changes at the protein level. Proteome quantification can be carried out on a relative or an absolute scale. Relative protein quantification methods like iTRAQ [1], SILAC [2] and label-free quantification [3,4] allow relative protein abundances to be compared in samples and the proteome dynamics to be characterized in cellular systems. However, absolute intracellular protein concentrations at the proteome level are essential for a quantitative and comprehensive understanding of an organism's metabolism and for mathematical modeling in systems biology. For instance, knowing intracellular protein concentrations enables one to evaluate the cost of running an active metabolic pathway, expressing enzymes for stress responses, estimate ribosomal translational capacity, etc.

Achieving absolute quantification of the whole proteome can be very expensive and laborious by using precise isotope dilution based methods like stable isotope labeled peptides [5,6] or

* *Corresponding author at:* Competence Center of Food and Fermentation Technologies, Akadeemia tee 15A, 12618 Tallinn, Estonia. Tel.: +372 55635885; fax: +372 6408282.
E-mail address: liisa@tftak.eu (L. Arike).

1874-3919/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.jprot.2012.06.020

proteins [7]. Absolute quantification can be determined using label-free quantification, which is cheaper and easier to perform; however, it is a less accurate alternative to isotope dilution based methods. In addition, several other features of label-free quantification should be considered, mainly related to sample handling/processing and LC-MS/MS analysis (protein extraction and digestion efficiency [8], ion suppression during ionization [9,10], reproducibility of retention times [11], etc.). Limitations and concerns of label-free quantification methods are reviewed thoroughly elsewhere [12–14].

Based on the quantification algorithms used, label-free methods can be divided into two classes: 1) those based on the measurement of precursor ion current areas (e.g. MS^E [15], T3PQ [16], iBAQ [17]), or 2) those based on tandem MS data, e.g. protein sequence coverage or spectral counting (e.g. emPAI [18] and APEX [19]). Protein absolute abundances can be determined by label-free methods with or without standards. The more cost-efficient and easier option is to exclude the standard proteins and calculate protein abundances from the fraction of each protein in the total protein pool. This has been done with APEX [19], emPAI [18,20] and iBAQ [21] methods, assuming that most of the proteins which contribute to the total protein pool are identified and quantified. Quantification accuracy can be increased by using a standard curve from a mixture of proteins with known amounts, differing in size and concentration [15,17].

To date, several studies have been performed to compare different label-free absolute quantification methods with each other [16,22] (for relative scale comparison see also a review [12]), with some combining the alternative approaches to gain from strengths of different methods [23,24]. A comparison of two-dimensional gel electrophoresis and APEX for absolute proteome quantification in *Shigella dysenteriae* cells showed a reasonably good correlation (R^2=0.67) for 255 protein quantities determined by both methods [22]. Spectral counting methods APEX and emPAI have been compared previously to the precursor signal intensity based method [16]. It was demonstrated with samples containing different amounts of four standard proteins or yeast extract spiked with fetuin that for higher protein concentrations or more complex samples the spectral counting methods suffer from saturation effects. It was also found that the variance among the three replicates is surprisingly large for the spectral counting methods while the calculation method of the three most intense tryptic peptides peak area (T3PQ) is more reproducible. However, correlation of spectral counting and peak area calculating methods was not evaluated on a larger scale.

Malmström et al. combined the accuracy of spiked-in isotope-labeled standards with the dynamic range and coverage of label-free shotgun quantification to estimate proteome-wide protein copy numbers per cell [23]. Cellular concentrations of 769 proteins (accurate to ~2-fold changes) were estimated using the precursor ion intensity of the three most intense tryptic peptides [15]. In addition, spectral counting (APEX method [19]) was used to cover low abundance proteins with an accuracy of ~3-fold for 1095 additional proteins. Correlation between spectral counting and extracted precursor ion intensities to absolute abundance data was found to be good on a logarithmic scale for standard proteins (R^2=0.56, and R^2=0.86, respectively). No correlation between spectral counting and precursor ion intensities was reported.

In the current study, three different label-free approaches were used in order to calculate intracellular protein concentrations for every quantified protein. Spectral counting methods emPAI [18] and APEX [19] were chosen mainly because of the possibility to apply them on already existing data. The exponentially modified protein abundance index (emPAI) is an approximate protein quantification method based on experimentally observed peptides and the calculated number of observable peptides. Since emPAI is implemented in the Mascot database search platform, it is easy to use and the method is MS instrument independent. While emPAI employs unique peptide counts, the absolute protein expression (APEX) method uses redundant peptide counts and also the correction factor Oi, which estimates the number of expected unique peptides, calculated from their probability of being observed. Calculation of APEX values has been made very easy by the APEX Quantitative Proteomics Tool [25]. Thirdly, we chose a peak intensity-based absolute quantification method iBAQ [17]. Since iBAQ is integrated into the quantitative proteomics software package MaxQuant [26], which is capable of processing SILAC and label-free data, it offers new opportunities for data analysis whereby combinations of absolute and relative quantifications are easy to perform.

In this work we evaluate existing label-free methods in order to obtain absolute quantification of *Escherichia coli* proteome and analyze the obtained results from the perspective of cell physiology. We show that label-free quantification reasonably estimates protein abundances in the cell, producing proteome level data needed in systems biology for a comprehensive understanding of cell metabolism. Naturally, internal isotope-labeled standards should be used if more accurate concentrations of a few targeted proteins are required. We feel it is very important for the proteomics and systems biology communities to evaluate different absolute proteome quantification methods in terms of cost and data processing time without neglecting the physiological relevance of the quantitative data.

2. Materials and methods

2.1. *Sample preparation for a label-free experiment*

E. coli K-12 MG1655 (λ-, F-, *rph*-1, *Fnr*+; Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ), DSM No. 18039) was cultivated on glucose minimal medium in chemostat culture at a specific growth rate of 0.11 h^{-1} under the following conditions: temperature 37 °C, pH 7, agitation speed of 800 rpm, and aerobic conditions (air flow rate 150 ml/min). Three independent cultivation experiments were performed, which have been described in detail [27]. Proteome analysis was conducted in two independent biological experiments.

E. coli steady state chemostat culture was harvested in a 1 ml volume, washed once with PBS and flash frozen with liquid nitrogen until further processing. For cell lysis, cell pellets were suspended on ice in 200 μl urea lysis buffer (6 M urea/2 M thiourea in 10 mM Hepes, pH 8.0). Cells were disrupted with agitation using 100 mg glass beads at 4 °C for 15 min. Unbroken cells and glass beads were pelleted for 15 min at 4 °C at 14,800 rpm in a table-top centrifuge. Protein concentrations were determined with a 2D Quant kit (Amersham Biosciences, USA).

For proteome analysis 3 μg of protein was reduced for 30 min at room temperature with 10 mM dithiothreitol (DTT), followed by alkylation for 20 min with 50 mM iodoacetamide (IAA) in the dark at room temperature. Initial digestion was performed with endoproteinase LysC (Wako Chemicals USA, VA) at an enzyme to protein ratio 1:50 for 3 h at room temperature. Samples were then diluted four-fold with a digestion buffer, 50 mM aqueous ammonium and sequencing grade modified trypsin (Promega, Madison WI, USA) was added in enzyme to protein ratio 1:50. Samples were incubated overnight at room temperature. Trypsin and LysC activity was quenched by 0.1% trifluoroacetic acid (TFA) and peptides were desalted using C18-StageTips [28].

2.2. HPLC and mass-spectrometry

Peptides were analyzed in three technical replicates in order to improve the proteome coverage. LC-MS/MS analysis was performed using an Agilent 1200 series nanoflow system (Agilent Technologies) connected to a LTQ Orbitrap mass-spectrometer (Thermo Electron, San Jose, CA, USA) equipped with a nano-electrospray ion source (Proxeon, Odense, Denmark). Purified peptides were loaded on a self-packed fused silica emitter (150 mm × 0.075 mm, New Objective) packed with Reprosil-Pur C18-AQ 3 μm particles (Dr. Maisch, Germany) at a flow rate of 0.7 μl/min. Peptides were separated with a 240-minute gradient from 2 to 40% B (A: 0.5% acetic acid, B: 0.5% acetic acid/80% acetonitrile) using a flow-rate of 200 nl/min and sprayed directly into an LTQ Orbitrap mass-spectrometer (Thermo Electron, Germany) operated at 180 °C capillary temperature and 2.2 kV spray voltage.

Full mass spectra were acquired in a profile mode, with a mass range from m/z 300 to 1900 at a resolving power of 60,000 (FWHM). Up to five data-dependent MS/MS spectra were acquired in a centroid mode in the linear ion trap for each FTMS full-scan spectrum (normalized collision energy 35%, max injection time 150 ms, fill value 5×10^3). Each fragmented ion was dynamically excluded for 60 s.

The data associated with this manuscript may be downloaded from the ProteomeCommons.org Tranche network using the following hash: LZqGIJspJjapNouL1AVDPguIUOHZ66He5lCb/DhK9BI0k6USd02nSZWHAfrLttgNfxOnSn6Ha3VXMQBiX2Uj1+vICNEAAAAAAAAAFoQ==.

All quantified proteins and their properties used in the following analysis are listed in Supplementary Table 1.

2.3. Protein quantification

2.3.1. Intensity-based absolute quantification (iBAQ)

Intensity-based absolute quantification (iBAQ) was carried out as described elsewhere [17]. Briefly, the Universal Proteomics Standard (UPS2, Sigma-Aldrich), 10.6 μg was dissolved in a 20 μl of lysis buffer (7 M urea/2 M thiourea) and mixed with the protein sample prior to digestion as follows: 1.1 μg UPS2 + 3 μg *E. coli* lysate.

Raw data files were analyzed with the MaxQuant software package (version 1.1.0.36) [26]. Generated peak lists were searched using the Andromeda search engine (built into MaxQuant) against *E. coli* database (downloaded 12.07.2010 from http://cmr.jcvi.org/). The database was supplemented with UPS protein sequences as well as with common contaminants (e. g. human keratin, trypsin). MaxQuant searches were performed with full tryptic specificity, a maximum of two missed cleavages and a mass tolerance of 0.5 Da for fragment ions. Carbamidomethylation of cysteine was set as a fixed modification and methionine oxidation and protein N-terminal acetylation were set as variable modification. The required false discovery rate (FDR) was set to 1% both for peptide and protein levels and the minimum required peptide length was set to six amino acids. In addition, "Match between runs" option with a time window of 1.5 min was allowed, as was the iBAQ quantification option.

Protein copies per cell were calculated by multiplying the molar concentration with the Avogadro constant and dividing with the number of cells in the respective experiment obtained by plate counting ($8–9 \times 10^9$ cells/ml) [27].

2.3.2. Spectral counting based absolute quantification

2.3.2.1. Mascot search. Fragment MS/MS spectra from raw files were extracted as MSM files and then merged to peak lists using the Raw2MSM version 1.7 [29], selecting top six peaks for 100 Da. MSM files for the three technical replicates of the same sample were concatenated to generate a single large peak list file with a MultiRawPrepare.pl script (http://msquant.alwaysdata.net) and subsequently searched with the Mascot 2.2 search engine (Matrix Science, London, UK) against the *E. coli* K-12 MG1655 protein sequence database downloaded 22.09.2009 from EcoGene 2.0 (http://ecogene.org), supplemented with common contaminants. Search parameters were as follows: two missed trypsin cleavage, fixed modification was set as carbamidomethyl (C), variable modifications were set as oxidation (M) and acetyl (protein N-term), 5 ppm precursor mass tolerance and 0.6 Da MS/MS mass tolerance. In order to estimate the false discovery rate (FDR) decoy search option was allowed.

2.3.2.2. Absolute protein expression index (APEX). The Mascot search results were validated by the PeptideProphet and ProteinProphet algorithms [30] before the absolute protein expression indexes (APEX) [19] were calculated by the APEX Quantitative Proteomics Tool [25]. An estimated false positive rate (FPR) cut-off of less than 5% was used, which corresponded to the ProteinProphet probability $p > 0.5$. FPR < 5% was chosen, as this resulted in a reasonable number of quantified proteins (1220), comparable with the iBAQ dataset (1334 proteins). Limiting the FPR to less than 1% would result in the loss of more than 200 proteins.

Total concentration of protein copies per cell (at a specific growth rate of 0.11 h^{-1}) was calculated based on biomass concentration, determined gravimetrically as dry cellular weight described by Nahku et al. [31], and protein concentration measured using the Lowry method [32]. Taking into account the weighted average molecular mass of 1000 most abundant proteins and cell counts based on plate counts we estimated 2.3×10^6 protein copies per *E. coli* cell at a specific growth rate of 0.11 h^{-1}; the value of 2×10^6 protein copies per cell was used in further calculations, accounting for the 12% loss of protein due to insufficient cell lysis, as determined by the gap between the results obtained by the Lowry protein measurement and the 2D kit protein measurement in the cell lysate. Total protein copies per cell value were used as a

normalization factor to determine individual protein copies per cell values for all identified proteins from the APEX indexes. Additionally, UPS2 standard curve was used as in iBAQ calculations in order to compare the effect of different calculation methods.

2.3.2.3. The exponentially modified protein abundance index (emPAI). The exponentially modified protein abundance index (emPAI) [18] values were obtained directly from the Mascot database search. Data were filtered to the FDR threshold less than 1%. Protein copies per cell for each protein were calculated by dividing each individual protein emPAI value by the sum of all emPAI values and taking into account the total protein copies per cell value explained above (2×10^6 copies per cell) as a normalization factor. As with APEX, UPS2 standard curve was used as an alternative method to calculate cellular abundances for each quantified protein.

2.4. Data integration and analysis

Data from three different label-free absolute quantification experiments were merged together. All identified and quantified proteins were grouped into clusters of orthologous groups (COG) [33] and divided into transcription units and functional complexes according to the EcoCyc database [34] using the in-house script.

All the correlations reported are Pearson squared correlation coefficients of logarithmized values if not otherwise stated. Variability is characterized by the coefficient of variation (CV, %), which is defined as the ratio of the standard deviation to the arithmetic mean.

The codon adaptation index (CAI) values for each protein were calculated based on protein abundances in the iBAQ dataset as follows: proteins which mass accounted for more than 0.5% of the whole protein mass were chosen for codon usage table calculation with the EMBOSS online tool [35]; the acquired codon usage table was used to calculate CAI values using the Seqinr package in the R environment [36].

The cost of protein synthesis was calculated for all quantified proteins by multiplying the respective protein's abundance in the cell with its peptide bond count and 4.306 [37], which stands for the cost in ATP for one amino acid polymerization reaction in the ribosome.

Apparent enzyme activities (kcat) per protein chain or subunit (without taking into account the number of proteins and catalytic sites necessary for catalytic activity) were estimated by iBAQ abundances (protein copies in g-DCW) and the ratios of specific flux values in mmol g-DCW^{-1} h^{-1} (g-DCW — grams of dry cellular weight), using previously published values of specific fluxes obtained in the same experiments [27]. Apparent kcat calculations were based on absolute amounts of 190 enzymes and 60 metabolic fluxes in the main metabolic network; covering glycolysis, tricarboxylic acid cycle, pentose phosphate pathway, respiratory chain and biopolymer monomer synthesis (see Supplementary Table 1).

Staircase-like expression was analyzed for transcription units with two or more components. Protein expression levels of transcription units were sub-divided into "no staircase" and staircase-like behavior types "up," "down" and "others." A transcription unit was classified as "no staircase" if at least half of its consecutive genes were not differentially expressed. Two consecutive genes in the transcription unit were considered differentially expressed if their protein abundance measurements for two biological replicates did not overlap. Staircase-like behavior expression of transcription units was classified as "up" or "down" if at least half of its consecutive genes were differentially expressed at higher or lower levels, respectively, in the mRNA emerging direction during transcription (5′→3′). The remaining transcription units were classified as "others."

3. Results and discussion

3.1. Comparison of different label-free quantification methods

3.1.1. Validation of quantification approaches

Absolute protein abundances can be calculated by normalizing individual protein contributions to the total protein mass in the sample. However, this method is dependent on the measured total protein amount and on the number of identified proteins. Another approach would be to add a non-labeled internal standard mixture, which consists of proteins that are different from those present in the sample. Therefore, we first decided to investigate the effect of internal standard addition on the performance of label-free quantification methods. For that, we included (according to the intensity-based absolute quantification (iBAQ) protocol [17]) the Universal Proteomics Standard (UPS2, Sigma Aldrich), which is a mixture of 48 precisely quantified human proteins with a dynamic concentrations range spanning five orders of magnitude.

Two different approaches were used to calculate the absolute protein concentration: 1) using linear relationship of a standard curve based on the known amounts of spike-in standard proteins (UPS2, Sigma); 2) normalizing individual protein contributions to the amount of protein analyzed. Comparison of standard protein abundances calculated by the latter two approaches revealed no difference in the Pearson squared correlations for the spectral counting methods APEX and emPAI (Supplementary Fig. 1B and C vs. E and F), while correlation decreased from 0.94 to 0.92 for the iBAQ with the normalization method (Supplementary Fig. 1A and D). Relying on the dynamic range and linear regression of the calibration curves, the iBAQ method using internal standards performed the best: its dynamic range spanned four orders of magnitude with $R^2=0.94$ compared to the dynamic range covering three orders of magnitude and $R^2=0.88$ for APEX and 0.83 for emPAI (Fig. 1A–C). Recent studies of spectral counting methods have demonstrated that optimal MS configurations are crucial in order to maximize the number of low abundant proteins quantified while keeping the estimates for the highly abundant proteins within the linear dynamic range [38]. Dynamic exclusion (DE) parameters can have a significant impact on the peptides and spectral counts detected and identified — two studies have determined optimal DE setting 90 s [38,39]. A 60-second dynamic exclusion was used in the current study, which may affect the dynamic range and quantification of low abundant proteins.

Internal standard enables to evaluate the magnitude of absolute protein abundances: the sum of all proteins in a cell

according to iBAQ and emPAI was 8 and 5% less than the value derived from the Lowry total protein analysis. This very small difference between the total protein amount measured by the colorimetric assay and the label-free quantitative proteomics methods (iBAQ and emPAI) indicates high confidence of our protein abundance datasets. Interestingly, the APEX method overestimates total protein concentration 1.5 times compared to the Lowry, iBAQ and emPAI methods.

After comparing the correlation of biological replicates (Supplementary Fig. 1) and equimolarity between ribosomal proteins (Supplementary Fig. 2) we concluded that the normalization approach suited better for the spectral counting methods APEX and emPAI in terms of squared Pearson correlation and the coefficient of variation (CV) than using internal standards. However, for the peak intensity based method iBAQ, quantification using calibration curve proved to be more accurate. Therefore, we decided to quantify protein abundances using the most appropriate approach for each method: normalization for the APEX and emPAI, and the internal standard calibration for the iBAQ.

3.1.2. *Comparison of protein abundances calculated by different label-free quantification methods*

Correlation between biological replicates for the iBAQ was found to be 0.92 with an average CV of 15% (Fig. 1D) while the APEX and emPAI performed slightly worse: R^2=0.83 and CV=23% (Fig. 1E), R^2=0.92 and CV=20% (Fig. 1F), respectively. This high correlation between biological replicates can be accounted for by the strictly controlled cell cultivation systems used in this study, for which high reproducibility of biomass yield, product consumption and formation rates and also gene expression levels have been reported by us earlier for *E. coli* K-12 MG1655 [27,31] and *Lactococcus lactis* IL1403 [40,41].

We found good correlation between protein abundances determined by different absolute label-free quantification methods. The spectral counting method APEX versus the peak intensity measurement method iBAQ resulted in R^2=0.76 (Fig. 1G) and the correlation between other spectral counting methods emPAI and iBAQ was found to be 0.81 (Fig. 1H). Correlation between the two spectral counting methods emPAI and APEX was found to be 0.77 (Fig. 1I).

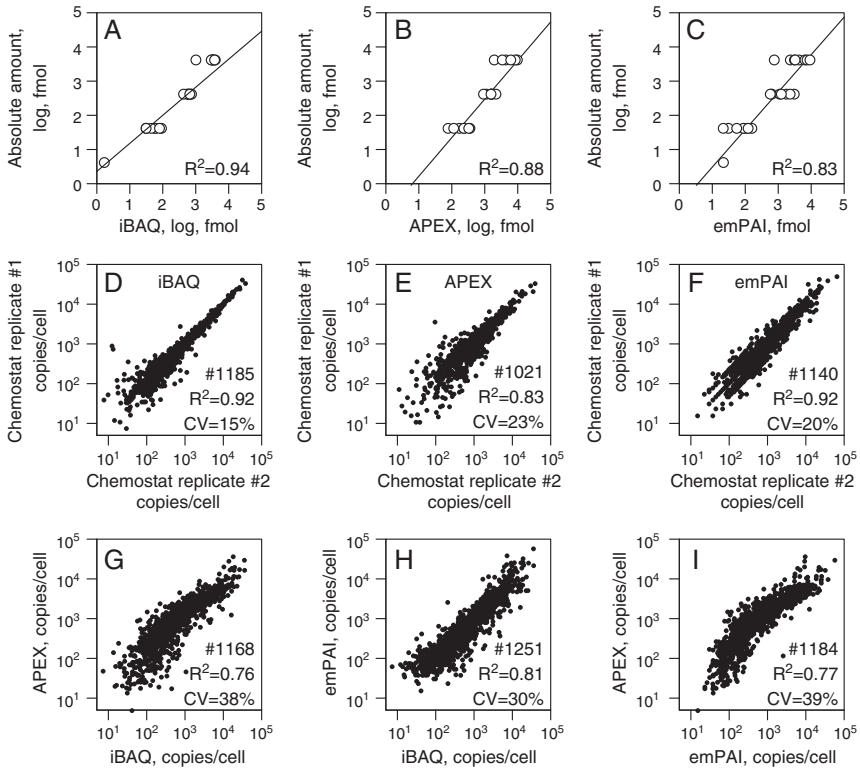


Fig. 1 – Comparison of absolute abundances obtained by different label-free quantification methods. All data are in logarithmic scale, R^2 — squared Pearson product moment correlation; # — number of quantified proteins; CV — correlation of variation in percentage. When different methods were compared, also proteins quantified only in one replicate were included. A–C) Correlation between absolute amounts of UPS2 standard proteins and values calculated by different approaches of label-free quantification methods. D–F) Correlation of replicate chemostat experiments. G) Correlation of spectral counting method APEX to intensity based method iBAQ. H) Correlation of spectral counting method emPAI to intensity based method iBAQ. I) Correlation of spectral counting methods emPAI and APEX.

All the quantification methods were based at least on one peptide quantification, since when limiting quantification to the most commonly used requirement of two peptides, no substantial improvements to the quality of the dataset were seen. The R^2 between biological replicates did not improve significantly (Fig. 1 vs. Supplementary Fig. 3), and the average CV between biological replicates improved only by some percent (15 to 13% in the case of the iBAQ method, 23% vs. 21% in the case of the APEX; 20% vs. 19% in the case of the emPAI, Fig. 1 and Supplementary Fig. 3). If quantifications were limited to at least two peptides per protein, more than a hundred mainly low concentration proteins could not be quantified (22, 10 and 6% of all proteins for emPAI, iBAQ and APEX, respectively).

3.1.3. *Absolute proteome comparison with published E. coli datasets*

To the best of our knowledge, our iBAQ proteome dataset is the first for *E. coli*; thus, no comparison could be carried out in this case. The APEX quantification method was originally demonstrated for proteome characterization of yeast, *E. coli* and mouse T-cell lymphoma cells [19]. We compared our data to the 449 proteins quantified by the APEX method in *E. coli* strain K-12N3433 reported by Lu et al. [19], a dataset containing 600 proteins less than ours, mostly lacking proteins from the lower abundance range (Fig. 2A). Comparison of their data with our iBAQ, APEX and emPAI values yielded squared Pearson correlation coefficients of 0.47, 0.36 and 0.40, respectively (Supplementary Fig. 4A–C). Since APEX quantification is influenced by the correction factor Oi, the low correlation between our APEX and previously published data [19] could be explained by different Oi values; however, the Pearson squared correlation between Oi values was found to be very high — 0.97 (data not shown).

Ishii et al. measured absolute values for 52 enzymes by using isotope-labeled proteins in *E. coli* K-12 BW25113 chemostat culture at a specific growth rate of 0.1 h^{-1} [42]. Comparison of these absolutely quantified 52 enzymes and our data resulted in moderate Pearson correlation coefficients 0.51, 0.42 and 0.40 for iBAQ, APEX and emPAI, respectively (Supplementary Fig. 4D–F).

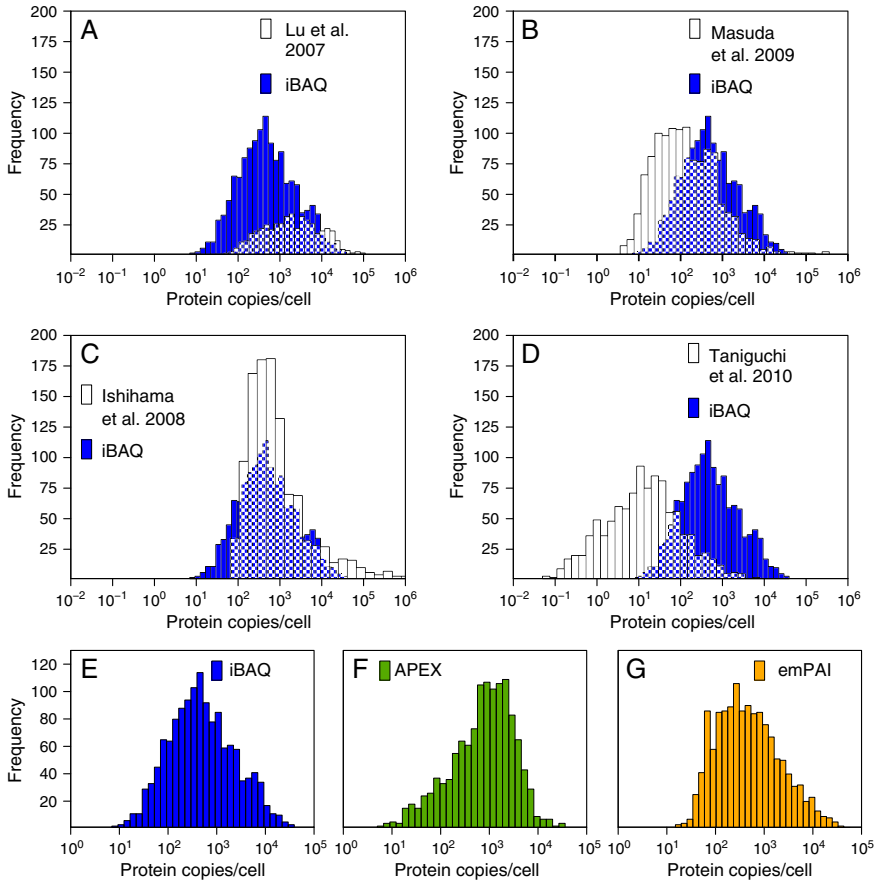


Fig. 2 – Comparison of protein abundance distribution of the current study and published data for *E. coli* proteome. Histograms represent protein cellular abundance distributed to 30 bins. A) Protein abundances calculated using the APEX method [19]. B–C) Protein abundances calculated using the emPAI method [20,43]. D) Protein abundances measured using a yellow fluorescent protein fusion library [44]. E–G) Protein abundance distribution for datasets calculated in the current study.

Moderate correlation (R^2=0.3–0.41) was also found between our data and those of Ishihama et al. [20] and Masuda et al. [43] who used the emPAI method in case of *E. coli* strains MC4100 and K-12 BW25113, respectively (Fig. 2B, C; Supplementary Fig. 4G–L). The dynamic range of the protein abundances observed in the current study was most comparable with the data of Masuda et al. (Fig. 2B) [43].

Ishihama et al. [20] detected Pearson correlation coefficient 0.84 with a p-value of $<10^{-10}$ (logarithmized variables; R^2= 0.71) when comparing their emPAI values with the 52 enzyme abundances determined using isotope-labeled proteins by Ishii et al. [42]. Comparing abundances of these 52 enzymes in our datasets with the emPAI values obtained by Ishihama et al. [20] revealed Pearson squared correlation coefficients of 0.44, 0.52 and 0.37 for iBAQ, APEX and emPAI, respectively (Supplementary Fig. 4G–L). As almost all correlations were in the same range we concluded that overall moderate correlations between different studies may be explained mainly by the fact that different *E. coli* strains and growth conditions were used in those studies.

Lately, protein abundances in *E. coli* at single cell level were measured by yellow fluorescent protein fusion library [44]. We detected poor correlation between that and our datasets: 0.17, 0.17 and 0.14 with iBAQ, APEX and emPAI, respectively (Supplementary Fig. 4M–O). Notably, Taniguchi et al. quantitative values differ in average by two orders of magnitude from our data (Fig. 2D) and the abundances translate into a hundred times lower total protein amount in the cell compared to our results. This is also in accordance with their comparison with Lu et al. — almost hundred times difference in protein abundances between the two studies was observed [44].

3.2. *Escherichia coli* proteome abundance analysis

Median protein copy numbers per cell were 457, 886 and 409 and for iBAQ, APEX and emPAI, respectively. We found that the top 20% of proteins by abundance contributed to 76%, 62% and 78% of the total protein amount in the cell for iBAQ, APEX and emPAI, respectively. This is in accordance with the well-known understanding that a small fraction of proteins are of high abundance, quantitatively presented in several studies for mammalian cells [21,45], *Saccharomyces cerevisiae* [46], *Leptospira interrogans* [24] and *Mycoplasma pneumoniae* [47].

Proteome data are not expected to be normally distributed based on studies of yeast [48], while studies of *E. coli* have presented a normal distribution [19,44]. Interestingly, in our case protein abundance distribution seems to depend on the quantification method used: cellular protein abundances were normally distributed for iBAQ while results tended to cluster more towards high and low abundance proteins for APEX and emPAI, respectively (Fig. 2E–G).

3.2.1. *Protein abundance compared to protein length and CAI*

It has been speculated that smaller proteins are present in the cell at higher copy numbers compared to larger ones as a way to minimize transcriptional and translational costs [49]. Although we found low correlation between protein abundances and protein length, highly abundant proteins tend to be shorter (Supplementary Fig. 5A–C), similarly as noticed by Ishihama et al. [20] and Schwanhäusser et al. [17].

Translation efficiency of genes can be described by codon usage bias for which the codon adaption index (CAI) [50] is a good measure. We found R^2 of around 0.19–0.27 between protein copy numbers and CAI (Supplementary Fig. 5D–F), depending on the quantification method used. Lu et al. reported for *E. coli* low correlation R^2=0.33 between protein abundance and CAI [19]. Ishihama et al. reported R=0.57 (R^2=0.32) between log-copy number and CAI for *E. coli* [20]. We also noted that proteins with smaller CAI are less frequently indentified and quantified, and proteins with CAI smaller than 0.16 were not identified and quantified at all (Supplementary Fig. 5G).

3.2.2. *Proteome coverage and distribution*

Next, we analyzed our proteome datasets in more detail by grouping all the quantified proteins to the clusters of orthologous groups (COG) functional classes (Fig. 3A). Overall, the different label-free quantification methods showed similar results for COG protein abundance percentages in the total protein pool. The only significant difference was detected for group J, which embraces proteins involved in translation, ribosomal structure and biogenesis. The highest cellular abundance for group J was detected with emPAI and the lowest with APEX methods (Fig. 3A), mostly due to high differences of abundance for ribosomal proteins (Fig. 4C) and elongation factor EF-Tu (57,072 and 29,430 copes/cell with emPAI and APEX methods, respectively). Since all three methods performed similarly and iBAQ presented the smallest CVs and the best correlation for biological replicates, the following discussion is mainly concentrated on iBAQ data.

Proteome coverage of *E. coli* 4333 protein encoding genes [51] was 31% and proteome coverage of the COG classes was in average 34% (Supplementary Fig. 6). Most abundant group J made up 21% of the total protein cellular abundance (Fig. 3A) and had the best COG coverage of 70% (Supplementary Fig. 6). Protein groups involved in energy production and conversion (C), carbohydrate transport and metabolism (G), amino acid transport and metabolism (E) showed also high cellular abundances, 10%, 11% and 9%, respectively (Fig. 3A), with a COG coverage of 37% 34% and 42%, respectively (Supplementary Fig. 6).

Protein synthesis, or more specifically the polymerization of amino acids, is the largest energy-consuming process in the cell, with more than 45% of the overall ATP consumption [37]. In order to comprehend the metabolic burden of protein synthesis for each COG class, we calculated the cost of expression of proteins in molecules of ATP (see Materials and methods for calculation details). This analysis yielded an interesting result: although the group J (translation, ribosomal structure and biogenesis proteins) showed the highest percentage from the total protein pool, the cost for expressing group C (proteins involved in energy production and conversion) was the highest: 15% for group J compared to 17% for group C (Fig. 3B). The cost for expression of proteins involved in carbohydrate, amino acid and nucleotide transport and metabolism (G, E and F) was altogether 28%, which indicates that the metabolic burden for protein synthesis is higher for biomass formation than that for energy formation.

Apparent kcat values per protein chain or subunit were calculated in order to estimate enzyme activities without *in vivo* assays (see Materials and methods for calculation details) (Fig. 3C). We found that biosynthetic enzymes (COG functional classes G, E, F) are working with ten times lower

activity (median <10 s^{-1}) than energy generating enzymes (COG functional class C; median is 63 s^{-1}). High enzymatic activity for energy generating enzymes indicates a shortage of such genes, which can be a limiting factor for biomass or product formation rate.

3.2.3. *Protein organization into transcription units*

As most of the bacterial genes are organized to polycistronic transcription units [52], similar absolute abundances could be expected for proteins within transcription units. The average CV of protein abundance for 251, 274 and 277 transcription units with APEX, iBAQ and emPAI was 53%, 60% and 60%, respectively. The CV over all quantified proteins (1021, 1185 and 1140) was found to be 157%, 205% and 237% for APEX, iBAQ and emPAI, respectively, which is more than three times higher than within the transcription units (Supplementary Fig. 7A). This is in accordance with a previous study for operons of *L. interrogans* proteome [24]. Ishihama et al. noted an abundance variance within most of the operons studied being smaller than variance over all proteins [20].

A recent genome-wide transcriptomics study has revealed that in *M. pneumoniae* consecutive genes within the operons do not have the same expression level, leading to operon polarity [53], almost half of the 139 polycistronic operons showed staircase-like decay behavior, following a 5′ to 3′ direction. However, Schmidt et al. discovered staircase behavior on proteome level for only a minority of *L. interrogans* operons (~5%) [24]. *E. coli* has been shown to express many alternative transcripts within operons under various growth conditions [54]; therefore, to exclude artificial staircase behavior of protein

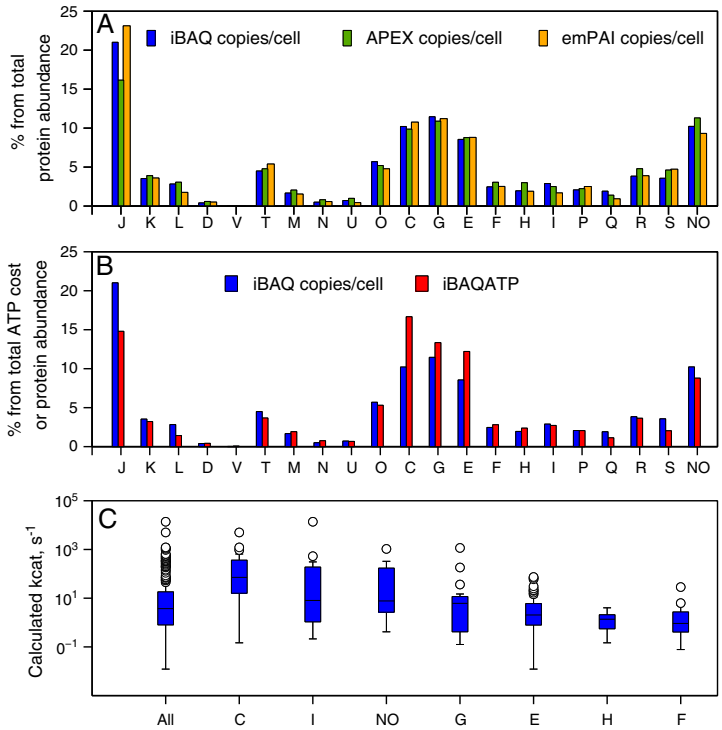


Fig. 3 – Analysis of the Clusters of Orthologous Groups (COG). A–B) Comparison of distribution to COG classes according to abundance in cells (A), and according to cost in ATP (B). C) Box plots showing distribution of catalytic activities of enzymes divided into COGs. Horizontal bars represent 25th, 50th (median) and 75th percentiles and whiskers represent 1.5 interquartile ranges. Outliers are plotted individually in open circles. J — Translation, ribosomal structure and biogenesis; K — transcription; L — replication, recombination and repair; D — cell cycle control, cell division, chromosome partitioning; V — defense mechanisms; T — signal transduction mechanisms; M — cell wall/membrane/envelope biogenesis; N — cell motility; U — intracellular trafficking, secretion, and vesicular transport; O — posttranslational modification, protein turnover, chaperones; C — energy production and conversion; G — carbohydrate transport and metabolism; E — amino acid transport and metabolism; F — nucleotide transport and metabolism; H — coenzyme transport and metabolism; I — lipid transport and metabolism; P — inorganic ion transport and metabolism; Q — secondary metabolites biosynthesis, transport and catabolism; R — general function prediction only; S — function unknown; NO — no COG class.

expression at the operon level we focused our analysis on transcription units to study protein expression behavior at the smallest transcription scale. Transcription unit protein expression levels were sub-divided into "no staircase" and staircase-like behavior types "up," "down" and "others." Our analysis revealed a very high presence (91%, 88% and 87% for iBAQ, APEX and emPAI, respectively) of transcription units with staircase-like protein expression (Supplementary Fig. 8). Further subdividing of the polycistronic transcription units by their staircase behavior and number of members showed similar distribution between groups of "up," "down" and "others" (Supplementary Fig. 8). The existence and high percentage (33%) of "up-like" staircase behavior were surprising since these have not been observed at mRNA level [53]. Additional analysis with mRNA abundances in the future should shed light into the possibility of a compensatory mechanism between mRNA and protein expression patterns within transcription units.

3.2.4. *Protein complex abundances and stoichiometry*

Proteins organized into complexes play an important role in cellular metabolic functions as enzymes, chaperones, ribosomes or transport systems. *E. coli* protein complexes are relatively well covered and more than 270 multimer complexes can be found in the EcoCyc database at http://www.ecocyc.org [34]. It is a challenge to cover all the complexes by quantitative proteomics and we ended up analyzing 118 complexes where at least two components were quantified (Supplementary Table 2). Similarly to transcription units, we found a three times lower CV among the complexes compared to all quantified proteins. An average CV among the complexes was found to be 58%, 64% and 56% for APEX, iBAQ and emPAI, respectively (if the stoichiometry was considered) (Supplementary Fig. 7B).

Absolute protein abundances enable experimental ratios to be compared in complexes with stoichiometries known to exist from previous studies. We found high correlation between known and experimental stoichiometry for some well-known complexes. RNA polymerase core enzyme RpoA/RpoB/RpoC, with a theoretical ratio of 2:1:1, was found to have copies per cell ratio close to 4000:2000:2000 with all the methods (Fig. 4A); succinyl-CoA synthase SucC/SucD, with a theoretical ratio of 2:2, had a similar experimental ratio with all the quantification methods used (Fig. 4A) and synthetases

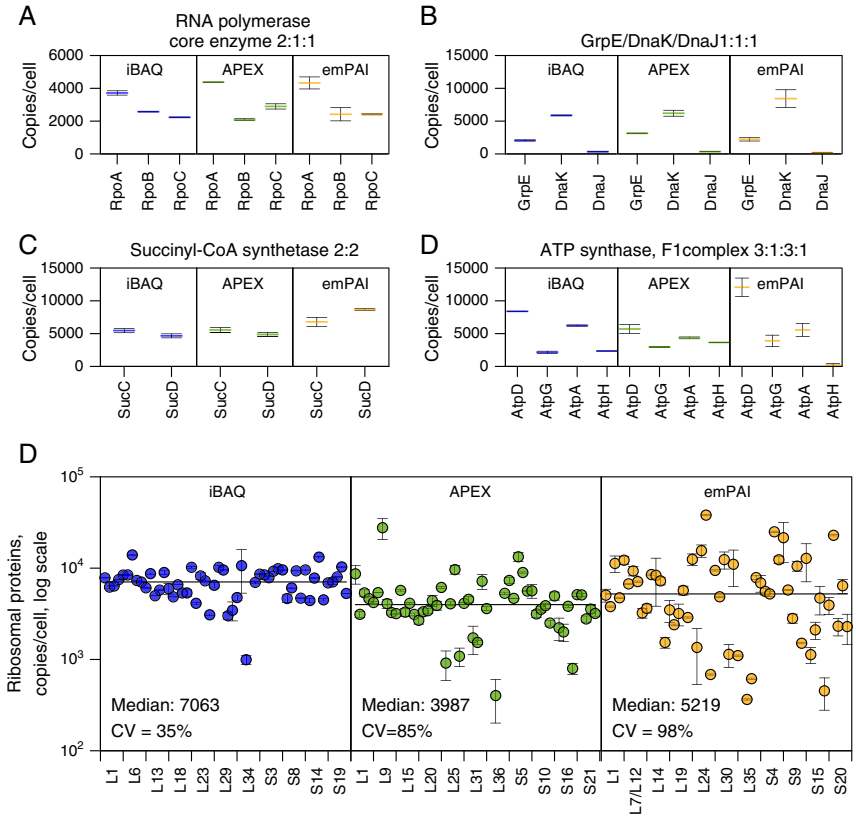


Fig. 4 – Stoichiometry of protein complexes. A) Examples of complexes which correlate with the known stoichiometry from the literature. Lines denote average and error bars absolute difference of two biological experiments. Expected stoichiometry for the complex is shown above the graph. B) Examples of complexes which do not correlate with the known stoichiometry from the literature. C) Abundance distribution of ribosomal proteins. Dots illustrate average and error bars absolute difference of two biological experiments. Dashed lines represent median ribosomal protein copy numbers.

RfbD/RfbC and GltD/GltB with 1:1 ratios had also very good correlation with theoretical stoichiometry (Supplementary Fig. 9A).

Experimental cellular protein abundances in this study did not mirror their functional stoichiometries for dynamic protein complexes, such as the sigma factors RpoDEFNS associated with RNA polymerase (Supplementary Fig. 9B) or the nucleotide exchange factor GrpE associated with the chaperones DnaK and DnaJ (Fig. 4B). The latter was also found by Maier et al. for *M. pneumoniae* [47]. Neither could we confirm as Maier et al. 1:1 stoichiometry for pyruvate dehydrogenase self-assembling complex LpD/AceF/AceE subunits AceF and AceE (Supplementary Fig. 9B). The third subunit protein Lpd is also shared with other complexes (2-oxoglutarate dehydrogenase and glycine cleavage multi-enzyme) and therefore has a more complicated stoichiometry. Kuntumalla et al. found lower APEX-calculated quantities than expected for ATP synthase F1 complex subunits AtpG and AtpH: ratio close to 8:1:8:1 instead of 3:1:3:1 for the AtpD/AtpG/AtpA/AtpH complex [22]. We noticed different results for all quantitative methods, with iBAQ data being closest to the theoretical results (Fig. 4B).

Complexes associated with membranes tend to have poor correlation with known stoichiometry, for example, succinate dehydrogenase SdhD/SdhB/SdhA and PTS transporters (Supplementary Fig. 9B). This phenomenon is most likely caused by the low solubility and loss of membrane proteins during the sample preparation [43].

Ribosomes are one of the largest protein complexes working in the cell and ribosomal proteins are expected to be expressed in equal copy numbers; however, we could not find agreement with the theoretical 1:1 stoichiometry. We identified and quantified 53 of 54 annotated ribosomal proteins and found that they span over one order of magnitude with the iBAQ method and over two orders of magnitude with spectral counting methods. Median ribosomal abundances were found to be 7063, 3987 and 5219 copies per cell with CVs of 35%, 85% and 98% for iBAQ, APEX and emPAI, respectively (Fig. 4C). This significant difference of variations among ribosomal protein abundances between peak area measurement and spectral counting methods shows that quantification methods have to be chosen with great care. In the literature, spectral counting has resulted in higher variations than peak area measurement using labeled peptides, probably due to the saturation effect in spectral counting for such high abundant proteins [20,47]. As ribosomal proteins are relatively short and have high lysine and arginine content, they produce a lot of tryptic peptides compared to their length, which can complicate label-free quantification of those proteins.

However, the ratio of ribosomal proteins is not clear, since also others have encountered problems to see equimolarity in ribosomal proteins. For instance, using the emPAI method, Ishihama et al. found that amounts of *E. coli* ribosomal proteins varied more than four orders of magnitude and did not correlate well with their detection frequencies, which indicate saturation effects [20]. Maier et al. quantified 43 ribosomal proteins in *M. pneumoniae* with labeled peptides and noticed cellular abundance differences of two orders of magnitude; they suggested that ribosomal proteins are not exclusively associated with the ribosome, instead they are present also as free monomers and some could be associated with different protein complexes [47].

4. Conclusion

We demonstrated by three label-free quantification methods that it is possible to obtain estimation of absolute protein abundances close to the realistic concentrations. Quantification based on peak intensity (iBAQ) was superior to the spectral counting methods (APEX, emPAI); however, all the used methods were able to produce similar information of *E. coli* proteome in terms of energy cost, distribution to COG classes and organization of proteins into transcription units or complexes.

We would like to encourage the generation and use of absolute quantitative proteome data, as it is essential for comprehensive understanding of the regulation mechanisms in the cell. Firstly, knowing the amount of proteins present in the cells allows us to rate energetic costs for several processes in the metabolism. Secondly, if flux values are added, apparent enzymatic activities can be estimated in order to understand the so-called “metabolic bottlenecks” in metabolism regulation, which could limit the overall rate of biomass or product formation. What is more, if the latter two levels would be accompanied also by mRNA absolute abundances, transcriptional/translational/post-translational regulation levels for each gene could be explained.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2012.06.020.

Acknowledgments

Financial support for this research was provided by the Enterprise Estonia project EU29994, and Estonian Ministry of Education and Research, through the grant SF0140090s08. The authors thank Klim Evdokimov and Andrus Seiman for providing data analysis scripts.

REFERENCES

[1] Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 2004;3:1154–69.
[2] Ong S-E, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, et al. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 2002;1:376–86.
[3] Chelius D, Bondarenko PV. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. J Proteome Res 2002:317–23.
[4] Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, et al. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem 2003;75:4818–26.
[5] Picotti P, Bodenmiller B, Mueller LLN, Domon B, Aebersold R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. Cell 2009;138:795–806.
[6] Picotti P, Rinner O, Stallmach R, Dautel F, Farrah T, Domon B, et al. High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nat Methods 2010;7:43–6.

[7] Hanke S, Besir H, Oesterhelt D. Absolute SILAC for accurate quantitation of proteins in complex mixtures down to the attomole level. J Proteome 2008;7:1118–30.
[8] Canas B, Pineiro C, Calvo E, Lopezferrer D, Gallardo J. Trends in sample preparation for classical and second generation proteomics. J Chromatogr A 2007;1153:235–58.
[9] Hirabayashi A, Ishimaru M, Manri N, Yokosuka T, Hanzawa H. Detection of potential ion suppression for peptide analysis in nanoflow liquid chromatography/mass spectrometry. Rapid Commun Mass Spectrom 2007;21:2860–6.
[10] King R, Bonfiglio R, Fernandez-Metzler C, Miller-Stein C, Olah T. Mechanistic investigation of ionization suppression in electrospray ionization. J Am Soc Mass Spectrom 2000;11:942–50.
[11] Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li G-Zhong, et al. Quantitative proteomic analysis by accurate mass retention time pairs. Anal Chem 2005;77:2187–200.
[12] Neilson KA, Ali NA, Muralidharan S, Mirzaei M, Mariani M, Assadourian G, et al. Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics 2011;11:535–53.
[13] Elliott MH, Smith DS, Parker CE, Borchers C. Current trends in quantitative proteomics. J Mass Spectrom 2009;44:1637–60.
[14] Zhu W, Smith JW, Huang C-M. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol 2010;2010:840518.
[15] Silva JC, Gorenstein MV, Li G-Z, Vissers JPC, Geromanos SJ. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. Mol Cell Proteomics 2006;5:144–56.
[16] Grossmann J, Roschitzki B, Panse C, Fortes C, Barkow-Oesterreicher S, Rutishauser D, et al. Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 2010;73:1740–6.
[17] Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, et al. Global quantification of mammalian gene expression control. Nature 2011;473:337–42.
[18] Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, et al. Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 2005;4:1265–72.
[19] Lu P, Vogel C, Wang R, Yao X, Marcotte EM. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 2007;25:117–24.
[20] Ishihama Y, Schmidt T, Rappsilber J, Mann M, Hartl FU, Kerner MJ, et al. Protein abundance profiling of the *Escherichia coli* cytosol. BMC Genomics 2008;9:102.
[21] Nagaraj N, Wisniewski JR, Geiger T, Cox J, Kircher M, Kelso J, et al. Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol 2011;7.
[22] Kuntumalla S, Braisted JC, Huang S-T, Parmar PP, Clark DJ, Alami H, et al. Comparison of two label-free global quantitation methods, APEX and 2D gel electrophoresis, applied to the *Shigella dysenteriae* proteome. Proteome Sci 2009;7:22.
[23] Malmström J, Beck M, Schmidt A, Lange V, Deutsch EW, Aebersold R. Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. Nature 2009;460:762–5.
[24] Schmidt A, Beck M, Malmström J, Lam H, Claassen M, Campbell D, et al. Absolute quantification of microbial proteomes at different states by directed mass spectrometry. Mol Syst Biol 2011;7:1–16.
[25] Braisted JC, Kuntumalla S, Vogel C, Marcotte EM, Rodrigues AR, Wang R, et al. The APEX Quantitative Proteomics Tool: generating protein quantitation estimates from LC-MS/MS proteomics results. BMC Bioinformatics 2008;9:529.
[26] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008;26:1367–72.
[27] Valgepea K, Adamberg K, Nahku R, Lahtvee P-J, Arike L, Vilu R. Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. BMC Syst Biol 2010;4:166.
[28] Rappsilber J, Mann M, Ishihama Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2007;2:1896–906.
[29] Olsen JV, de Godoy LMF, Li G, Macek B, Mortensen P, Pesch R, et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 2005;4:2010–21.
[30] Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 2003;75:4646–58.
[31] Nahku R, Valgepea K, Lahtvee P-J, Erm S, Abner K, Adamberg K, et al. Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures. J Biotechnol 2010;145:60–5.
[32] Lowry OH, Rosebrough NJ, Farr LA, Randall RJ. Protein measurement with the folin phenol reagent; 1951.
[33] Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated version includes eukaryotes. BMC Bioinformatics 2003;4:41.
[34] Keseler IM, Collado-Vides J, Santos-Zavaleta A, Peralta-Gil M, Gama-Castro S, Muñiz-Rascado L, et al. EcoCyc: a comprehensive database of *Escherichia coli* biology. Nucleic Acids Res 2011;39:D583–90.
[35] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000;16:276–7.
[36] Charif D, Thioulouse J, Lobry JR, Perrière G. Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics (Oxford, England) 2005;21:545–7.
[37] Stouthamer AH. A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie Van Leeuwenhoek 1973;39:545–65.
[38] Hoehenwarter W, Wienkoop S. Spectral counting robust on high mass accuracy mass spectrometers. Rapid Commun Mass Spectrom 2010;24:3609–14.
[39] Zhang Y, Wen Z, Washburn MP, Florens L. Effect of dynamic exclusion duration on spectral count based quantitative proteomics. Anal Chem 2009;81:6317–26.
[40] Adamberg K, Lahtvee P, Valgepea K. Quasi steady state growth of *Lactococcus lactis* in glucose-limited acceleration stat (A-stat) cultures. Antonie Van Leeuwenhoek 2009;95:219–26.
[41] Lahtvee P-J, Adamberg K, Arike L, Nahku R, Aller K, Vilu R. Multi-omics approach to study the growth efficiency and amino acid metabolism in *Lactococcus lactis* at various specific growth rates. Microb Cell Fact 2011;10:12.
[42] Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, et al. Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science 2007;316:593–7.
[43] Masuda T, Saito N, Tomita M, Ishihama Y. Unbiased quantitation of *Escherichia coli* membrane proteome using phase transfer surfactants. Mol Cell Proteomics 2009;8:2770–7.
[44] Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science (New York, NY) 2010;329:533–8.
[45] Beck M, Schmidt A, Malmström J, Claassen M, Ori A, Szymborska A, et al. The quantitative proteome of a human cell line. Mol Syst Biol 2011;7:1–8.
[46] Ghaemmaghami S, Huh W-K, Bower K, Howson RW, Belle A, Dephoure N, et al. Global analysis of protein expression in yeast. Nature 2003;425:737–41.

[47] Maier T, Schmidt A, Güell M, Kühner S, Gavin A-C, Aebersold R, et al. Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol Syst Biol 2011;7:1–12.
[48] Maier T, Güell M, Serrano L. Correlation of mRNA and protein in complex biological samples. FEBS Lett 2009;583:3966–73.
[49] Coghlan A, Wolfe KH. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. Yeast (Chichester, England) 2000;16:1131–45.
[50] Sharp PM, Li WH. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1987;15:1281–95.
[51] Riley M, Abe T, Arnaud MB, Berlyn MKB, Blattner FR, Chaudhuri RR, et al. *Escherichia coli* K-12: a cooperatively developed annotation snapshot—2005. Nucleic Acids Res 2006;34:1–9.
[52] Kozak M. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol Rev 1983;47:1–45.
[53] Güell M, van Noort V, Yus E, Chen W-H, Leigh-Bell J, Michalodimitrakis K, et al. Transcriptome complexity in a genome-reduced bacterium. Science (New York, NY) 2009;326:1268–71.
[54] Cho B-K, Zengler K, Qiu Y, Park YS, Knight EM, Barrett CL, et al. The transcription unit architecture of the *Escherichia coli* genome. Nat Biotechnol 2009;27:1043–9.

RECENT DISSERTATIONS DEFENDED AT TUT IN NATURAL AND EXACT SCIENCES

B71 **Cecilia Sarmiento.** *Suppressors of RNA Silencing in Plants.* 2008.
B72 **Vilja Mardla.** *Inhibition of Platelet Aggregation with Combination of Antiplatelet Agents.* 2008.
B73 **Maie Bachmann.** *Effect of Modulated Microwave Radiation on Human Resting Electroencephalographic Signal.* 2008.
B74 **Dan Hüvonen.** *Terahertz Spectroscopy of Low-Dimensional Spin Systems.* 2008.
B75 **Ly Villo.** *Stereoselective Chemoenzymatic Synthesis of Deoxy Sugar Esters Involving Candida antarctica Lipase B.* 2008.
B76 **Johan Anton.** *Technology of Integrated Photoelasticity for Residual Stress Measurement in Glass Articles of Axisymmetric Shape.* 2008.
B77 **Olga Volobujeva.** *SEM Study of Selenization of Different Thin Metallic Films.* 2008.
B78 **Artur Jõgi.** *Synthesis of 4'-Substituted 2,3'-dideoxynucleoside Analogues.* 2008.
B79 **Mario Kadastik.** *Doubly Charged Higgs Boson Decays and Implications on Neutrino Physics.* 2008.
B80 **Fernando Pérez-Caballero.** *Carbon Aerogels from 5-Methylresorcinol-Formaldehyde Gels.* 2008.
B81 **Sirje Vaask.** *The Comparability, Reproducibility and Validity of Estonian Food Consumption Surveys.* 2008.
B82 **Anna Menaker.** *Electrosynthesized Conducting Polymers, Polypyrrole and Poly(3,4-ethylenedioxythiophene), for Molecular Imprinting.* 2009.
B83 **Lauri Ilison.** *Solitons and Solitary Waves in Hierarchical Korteweg-de Vries Type Systems.* 2009.
B84 **Kaia Ernits.** *Study of In2S3 and ZnS Thin Films Deposited by Ultrasonic Spray Pyrolysis and Chemical Deposition.* 2009.
B85 **Veljo Sinivee.** *Portable Spectrometer for Ionizing Radiation "Gammamapper".* 2009.
B86 **Jüri Virkepu.** *On Lagrange Formalism for Lie Theory and Operadic Harmonic Oscillator in Low Dimensions.* 2009.
B87 **Marko Piirsoo.** *Deciphering Molecular Basis of Schwann Cell Development.* 2009.
B88 **Kati Helmja.** *Determination of Phenolic Compounds and Their Antioxidative Capability in Plant Extracts.* 2010.
B89 **Merike Sõmera.** *Sobemoviruses: Genomic Organization, Potential for Recombination and Necessity of P1 in Systemic Infection.* 2010.
B90 **Kristjan Laes.** *Preparation and Impedance Spectroscopy of Hybrid Structures Based on CuIn3Se5 Photoabsorber.* 2010.
B91 **Kristin Lippur.** *Asymmetric Synthesis of 2,2'-Bimorpholine and its 5,5'-Substituted Derivatives.* 2010.
B92 **Merike Luman.** *Dialysis Dose and Nutrition Assessment by an Optical Method.* 2010.
B93 **Mihhail Berezovski.** *Numerical Simulation of Wave Propagation in Heterogeneous and Microstructured Materials.* 2010.
B94 **Tamara Aid-Pavlidis.** *Structure and Regulation of BDNF Gene.* 2010.
B95 **Olga Bragina.** *The Role of Sonic Hedgehog Pathway in Neuro- and Tumorigenesis.* 2010.
B96 **Merle Randrüüt.** *Wave Propagation in Microstructured Solids: Solitary and Periodic Waves.* 2010.
B97 **Marju Laars.** *Asymmetric Organocatalytic Michael and Aldol Reactions Mediated by Cyclic Amines.* 2010.
B98 **Maarja Grossberg.** *Optical Properties of Multinary Semiconductor Compounds for Photovoltaic Applications.* 2010.
B99 **Alla Maloverjan.** *Vertebrate Homologues of Drosophila Fused Kinase and Their Role in Sonic Hedgehog Signalling Pathway.* 2010.
B100 **Priit Pruunsild.** *Neuronal Activity-Dependent Transcription Factors and Regulation of Human BDNF Gene.* 2010.
B101 **Tatjana Knjazeva.** *New Approaches in Capillary Electrophoresis for Separation and Study of Proteins.* 2011.
B102 **Atanas Katerski.** *Chemical Composition of Sprayed Copper Indium Disulfide Films for Nanostructured Solar Cells.* 2011.
B103 **Kristi Timmo.** *Formation of Properties of CuInSe2 and Cu2ZnSn(S,Se)4 Monograin Powders Synthesized in Molten KI.* 2011.

B104 **Kert Tamm.** *Wave Propagation and Interaction in Mindlin-Type Microstructured Solids: Numerical Simulation.* 2011.
B105 **Adrian Popp.** *Ordovician Proetid Trilobites in Baltoscandia and Germany.* 2011.
B106 **Ove Pärn.** *Sea Ice Deformation Events in the Gulf of Finland and This Impact on Shipping.* 2011.
B107 **Germo Väli.** *Numerical Experiments on Matter Transport in the Baltic Sea.* 2011.
B108 **Andrus Seiman.** *Point-of-Care Analyser Based on Capillary Electrophoresis.* 2011.
B109 **Olga Katargina.** *Tick-Borne Pathogens Circulating in Estonia (Tick-Borne Encephalitis Virus, Anaplasma phagocytophilum, Babesia Species): Their Prevalence and Genetic Characterization.* 2011.
B110 **Ingrid Sumeri.** *The Study of Probiotic Bacteria in Human Gastrointestinal Tract Simulator.* 2011.
B111 **Kairit Zovo.** *Functional Characterization of Cellular Copper Proteome.* 2011.
B112 **Natalja Makarytsheva.** *Analysis of Organic Species in Sediments and Soil by High Performance Separation Methods.* 2011.
B113 **Monika Mortimer.** *Evaluation of the Biological Effects of Engineered Nanoparticles on Unicellular Pro- and Eukaryotic Organisms.* 2011.
B114 **Kersti Tepp.** *Molecular System Bioenergetics of Cardiac Cells: Quantitative Analysis of Structure-Function Relationship.* 2011.
B115 **Anna-Liisa Peikolainen.** *Organic Aerogels Based on 5-Methylresorcinol.* 2011.
B116 **Leeli Amon.** *Palaeoecological Reconstruction of Late-Glacial Vegetation Dynamics in Eastern Baltic Area: A View Based on Plant Macrofossil Analysis.* 2011.
B117 **Tanel Peets.** *Dispersion Analysis of Wave Motion in Microstructured Solids.* 2011.
B118 **Liina Kaupmees.** *Selenization of Molybdenum as Contact Material in Solar Cells.* 2011.
B119 **Allan Olspert.** *Properties of VPg and Coat Protein of Sobemoviruses.* 2011.
B120 **Kadri Koppel.** *Food Category Appraisal Using Sensory Methods.* 2011.
B121 **Jelena Gorbatšova.** *Development of Methods for CE Analysis of Plant Phenolics and Vitamins.* 2011.
B122 **Karin Viipsi.** *Impact of EDTA and Humic Substances on the Removal of Cd and Zn from Aqueous Solutions by Apatite.* 2012.
B123 **David W. Schryer.** *Metabolic Flux Analysis of Compartmentalized Systems using Dynamic Isotopologue Modeling.* 2012.
B124 **Ardo Illaste.** *Analysis of Molecular Movements in Cardiac Myocytes.* 2012.
B125 **Indrek Reile.** *3-Alkylcyclopentane-1,2-Diones in Asymmetric Oxidation and Alkylation Reactions.* 2012.
B126 **Tatjana Tamberg.** *Some Classes of Finite 2-Groups and Their Endomorphism Semigroups.* 2012.
B127 **Taavi Liblik.** *Variability of Thermohaline Structure in the Gulf of Finland in Summer.* 2012.
B128 **Priidik Lagemaa.** *Operational Forecasting in Estonian Marine Waters.* 2012.
B129 **Andrei Errapart.** *Photoelastic Tomography in Linear and Non-linear Approximation* 2012.
B130 **Külliki Krabbi.** *Biochemical Diagnosis of Classical Galactosemia and Mucopolysaccharidoses in Estonia* 2012.
B131 **Kristel Kaseleht.** *Identification of Aroma Compounds in Food using SPME-GC/MS and GC-Olfactometry.* 2012.
B132 **Kristel Kodar.** *Immunoglobulin G Glycosylation Profiling in Patients with Gastric Cancer.* 2012.
B133 **Kai Rosin.** *Solar Radiation and Wind as Agents of the Formation of the Radiation Regime in Water Bodies.* 2012.
B134 **Ann Tiiman.** *Interactions of Alzheimer's Amyloid-Beta Peptides with Zn(II) and Cu(II) Ions.* 2012.
B135 **Olga Gavrilova.** *Application and Elaboration of Accounting Approaches for Sustainable Development.* 2012.
B136 **Olesja Bondarenko.** *Development of Bacterial Biosensors and Human Stem Cell-Based In Vitro Assays for the Toxicological Profiling of Synthetic Nanoparticles.* 2012.