Towards quantitative metabolomics and *in vivo* kinetic modeling in *S. cerevisiae*

André B. Canelas

Towards quantitative metabolomics and *in vivo* kinetic modeling in *S. cerevisiae*

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.dr.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties
in het openbaar te verdedigen op 28 juni 2010 om 12:30 uur

door

André DE CASTRO BIZARRO DUARTE CANELAS

Licenciado em Engenharia Biológica, I.S.T., U.T. Lisboa
geboren te Lisboa, Portugal

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr.ir. J.J. Heijnen
Copromotor: Dr. W.M. van Gulik.

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof.dr.ir. J.J. Heijnen	Technische Universiteit Delft, promotor
Dr. W.M. van Gulik	Technische Universiteit Delft, copromotor
Prof.dr. J.T. Pronk	Technische Universiteit Delft
Prof.dr. B. Teusink	Vrije Universiteit Amsterdam
Dr. B.M. Bakker	Rijksuniversiteit Groningen
Prof.dr.ing. M. Reuss	Universität Stuttgart
Dr.ir. W.A. van Winden	DSM Food Specialties
Prof.dr. P.D.E.M. Verhaert	Technische Universiteit Delft, reservelid

The studies described in this thesis were performed at the Bioprocess Technology Group of the Biotechnology Department at Delft University of Technology. The work was funded by SenterNovem through the IOP Genomics initiative, as part of the "Vertical Genomics: from gene to flux and back" project (IGE3006A), and carried out under the auspices of the Kluyver Centre for Genomics of Industrial Fermentation (NGI). The study described in Chapter 5 was carried out under the auspices of the Yeast Systems Biology Network (YSBN) coordination action (EC FP6 funding).
Cover art designed at www.tagxedo.com.
Printed in The Netherlands by CPI/WPS, Zutphen.

An experiment is a question which [the scientist] poses to Nature, and a measurement is the recording of Nature's answer.

Max Planck

Summary

The quantitative analysis of enzyme kinetics *in vivo*, leading to the construction of predictive kinetic models of metabolic reaction networks, is an old ambition in Biochemistry and a central goal within the emerging field of Systems Biology. However, this objective remains largely unfulfilled because of a fundamental problem: the gap between the overwhelming complexity of kinetic models and the limited availability and information content of *in vivo* data. To close this gap, efforts must be aimed at expanding our ability to generate high-quality quantitative *in vivo* data and at developing and improving approaches to manage the complexity of kinetic models. This thesis systematically addresses several of the key challenges towards bridging the gap between model and data, with the ultimate goal of enabling network-wide *in vivo* kinetic modeling, in *S. cerevisiae* as well as other biological systems.

Chapters 2 and 3 address the techniques used for sampling and sample treatment, which are crucial for ensuring the accuracy and precision of intracellular metabolite measurements. The issue of quenching yeast cells is analyzed in detail in **Chapter 2**. In particular, the question of whether leakage occurs during conventional cold methanol quenching, and to what extent, is still debated in the literature. Using a comprehensive mass balance approach we confirmed the occurrence of extensive leakage of metabolites from yeast cells. Furthermore, by varying the quenching conditions we successfully developed an improved method that entirely prevented leakage. The results also demonstrated quantitatively the need of the washing step to prevent overestimation of intracellular pools. In **Chapter 3** a procedure is proposed, based on differential additions of ^{13}C-labeled internal standards at different stages of sample treatment, which allows sensitive and comprehensive determination of metabolite recoveries with minimal experimental effort. This convenient approach should greatly facilitate method validation in the future. We applied the procedure for the validation of five different methods for the extraction of metabolites from yeast cells. The most striking observation was that one particular method (Freeze-Thawing in Methanol) which has been promoted by much of the recent literature in fact does not fulfill the most basic requirements. The findings in **Chapters 2 and 3** highlight the dangers of relying on insufficient method validation and the consequences that inadequate methodology can have for data quality.

Chapter 4 focuses on the issues of compartmentation and protein-binding, which can preclude the accurate determination of meaningful metabolite levels. We describe the successful application of the indicator reaction principle to determine the free NAD/NADH ratio in the cytosol of *S. cerevisiae*. The absence of adequate indicators in yeast was circumvented by expressing a heterologous enzyme (mannitol-1-phosphate 5-dehydrogenase), which allowed the cytosolic free NAD/NADH ratio to be calculated from the measured [fructose-6-phosphate]/[mannitol-1-phosphate] ratio. Remarkably, the estimated free cytosolic NAD/NADH values were more than one order of magnitude higher than the ratio determined from whole-cell total concentration of NAD and NADH. This difference

explained the often observed thermodynamic infeasibility problems in reactions involving NAD/NADH. It was also found that the estimated NAD/NADH ratio is extremely sensitive to changes in the external supply of electron donors and acceptors. These findings highlight the relevance of accurate compartment-specific data for understanding metabolic processes in eukaryotes.

Chapter 5 deals with the aspect of standardization of experimental and analytical methodologies, which is indispensable for enabling the interchange and integration of data from different studies/laboratories. Within a large collaborative effort, the "omics" platforms available at eight laboratories were compared in the analysis of a single set of biological experiments: a comparison of two yeast strains (the commonly used CEN.PK and a newly-constructed YSBN strain) under two growth conditions (batch and chemostat). Four different platforms used for analysis of transcripts (Affymetrix, Agilent, qPCR and TRAC) gave overall consistent results, although in one-on-one comparisons there were substantial discrepancies. The determination of metabolite concentrations by seven different labs resulted in differences of up to 3-fold, even for samples treated identically. Interestingly, metabolite ratios were more comparable and in terms of relative differences between strains or growth conditions, all platforms provided similar results. Determination of enzyme activities in two different labs gave comparable results. These results highlight the fact that obtaining fully reproducible absolute measurement between different labs is still a challenging task, even in the absence of biological/cultivation variation. Using the dataset generated, it was also possible perform an integrative comparison of two yeast strains, addressing in particular the differences observed in maximal growth rate and biomass yield. We found evidence of increased expression of both proteolitic and amino acid biosynthesis pathways in the CEN.PK strain, indicating that higher protein turnover leads to its lower biomass yield.

Finally, **Chapter 6** addresses the issue of managing the complexity of kinetic models, from a data-driven perspective. We introduce a thermodynamics-based approach that uses experimentally accessible *in vivo* data to categorize individual reactions, based on their thermodynamic state and kinetic behavior, into three categories: pseudo-, near- or far-from-equilibrium. We then demonstrate the approach by using the techniques developed and validated in **Chapters 2, 3 and 4** to generate a large high-quality metabolomics dataset comprising most of central metabolism, under a set of 32 conditions spanning flux ranges up to 60-fold. For 3/4 of the reactions analyzed we could obtain complete *in vivo*-derived mathematical descriptions which can be directly incorporated into a kinetic model. We also found that the *in vivo* kinetics of the reactions classified as near-equilibrium are dominated by the thermodynamic driving force and can be conveniently described as linear functions of the ratio between substrates and products (Q), a kinetic format which we dubbed "Q-linear kinetics". The reason why this simplification is possible could be traced back to the intrinsic correlation between changes in reactant concentrations that occur *in vivo*. We were also able, for the first time, to systematically estimate apparent *in vivo* Keq values. Remarkably, preliminary comparisons with data from *E. coli* suggested that they constitute a superior reference for precise *in vivo* thermodynamic analyses, compared to the respective *in vitro*-derived Keq data. These findings highlight the potential for data-driven model reduction by tailoring the complexity of rate expressions to the complexity of kinetic behavior displayed *in vivo*. The

approach proposed provides a framework for the initial stages of model formulation with minimal need for *a priori* information, which we believe will consitute the backbone for further kinetic modeling efforts.

Samenvatting

De kwantitatieve analyse van enzym kinetiek onder *in vivo* condities (de condities die heersen in de levende cel), onmisbaar voor het opstellen van voorspellende kinetische modellen van metabole reactienetwerken, is een oude ambitie in de biochemie en een centrale doelstelling binnen het opkomende gebied van de Systeembiologie. , Deze doelstelling blijft echter grotendeels onbereikbaar als gevolg van een fundamenteel probleem: de kloof tussen de enorme complexiteit van kinetische modellen en de beperkte beschikbaarheid van *in vivo* gegevens. Om deze kloof te dichten, zal meer aandacht moeten worden besteed aan het genereren kwantitatieve, onder *in vivo* condities verkregen data van hoge kwaliteit en op het ontwikkelen en verbeteren van methoden om de complexiteit van kinetische modellen binnen de perken te houden. Dit proefschrift behandelt, op systematische wijze, een aantal van de belangrijkste uitdagingen voor het overbruggen van de kloof tussen model complexiteit en data beschikbaarheid, met als einddoel het mogelijk maken van *in vivo* kinetische modellering van het gehele netwerk, in de gist Saccharomyces cerevisiae en andere biologische systemen.

Hoofdstukken 2 en 3 zijn gewijd aan de technieken die gebruikt worden voor bemonstering en monsterbehandeling, technieken die cruciaal zijn voor het waarborgen van de nauwkeurigheid en de precisie van intracellulaire metaboliet metingen. Bestaande methoden voor snelle fixatie (quenching) van het metabolisme van gistcellen worden in detail geanalyseerd in **Hoofdstuk 2**. De vraag of lekkage van metabolieten optreedt tijdens de conventionele quenching methode met koude methanol en zo ja in welke mate, is nog steeds onderwerp van discussie in de literatuur. Door het uitvoeren van metaboliet metingen in de diverse monster fracties en het opstellen van massabalansen werd aangetoond dat de lekkage vanuit de gistcellen significant is. Door het variëren van de quenching condities werd met succes een verbeterde quenching methode ontwikkeld, die lekkage volledig voorkomt. Aan de hand van de resultaten werd tevens kwantitatief aangetoond, dat een wasstap noodzakelijk is om overschatting van de intracellulaire metaboliet pools te voorkomen. In **hoofdstuk 3** wordt een procedure voorgesteld, gebaseerd op differentiële toevoeging van 13C-gelabelde interne standaarden in verschillende stadia van de monsterbehandeling, die een nauwkeurige bepaling van de metaboliet recovery mogelijk maakt met minimale experimentele inspanning. Deze aanpak leidde tot een aanzienlijke vereenvoudiging van de validatie van sampling en extractie methoden. Deze procedure werd toegepast voor de validatie van vijf verschillende methoden voor de extractie van metabolieten uit gistcellen. De opvallendste waarneming was dat één bepaalde methode (Freeze-Thawing in methanol) die wordt aanbevolen in een groot deel van de recente literatuur, in feite niet voldoet aan de meest elementaire eisen. De bevindingen in de

hoofdstukken 2 en 3 vestigen de aandacht op de gevaren van een vertrouwen op onvoldoende gevalideerde methoden en de gevolgen die gebrekkige methodologie kan hebben op de kwaliteit van de gegevens.

Hoofdstuk 4 richt zich op de problematiek van compartimentering en eiwitbinding, waardoor voor sommige metabolieten bepalingen van de gemiddelde concentratie voor de gehele cel van weinig betekenis zijn. In dit hoofdstuk wordt de succesvolle toepassing van het indicator reactie principe voor het bepalen van de vrije NAD/NADH-verhouding in het cytosol van *S. cerevisiae* beschreven. Het probleem dat een geen bruikbare indicator reactie voor dit doel in gist aanwezig was werd opgelost door de expressie van een heteroloog enzym (mannitol-1-fosfaat 5-dehydrogenase), waardoor de intracellulaire vrije NAD/NADH-verhouding kon worden berekend uit de gemeten [fructose-6-fosfaat]/[mannitol-1-fosfaat] ratio. Opmerkelijk is dat de op deze manier bepaalde vrije cytosolische NAD/NADH ratio's meer dan een orde van grootte hoger waren dan de ratio van de totale concentraties van NAD en NADH in de gehele cel. Dit verschil verklaart het vaak waargenomen probleem dat reacties waarbij NAD of NADH betrokken zijn vanuit thermodynamisch oogpunt niet zouden kunnen verlopen. Ook werd vastgesteld dat de op bovengenoemde wijze bepaalde vrije NAD/NADH-ratio zeer gevoelig is voor veranderingen in de externe aanvoer van elektrondonoren en -acceptoren. Deze bevindingen benadrukken het belang van nauwkeurige compartiment-specifieke gegevens voor het verkrijgen van een beter begrip van metabole processen in eukaryote organismen.

Hoofdstuk 5 behandelt het aspect van de standaardisatie van experimentele en analytische methoden, welke onmisbaar is om de uitwisseling en integratie van gegevens uit verschillende studiën en laboratoria mogelijk te maken. Binnen een groot Europees samenwerkingsverband, werden de "omics" platforms beschikbaar in acht laboratoria vergeleken door analyse van een aantal biologische experimenten: een vergelijking van twee giststammen (de meest gebruikte CEN.PK en een nieuw geconstrueerde YSBN stam) onder twee groeiomstandigheden (batch en chemostaat). Vier verschillende platforms voor de analyse van transcripten (Affymetrix, Agilent, qPCR en TRAC) gaven afzonderlijk consistente resultaten, hoewel in een-op-een-vergelijkingen er grote verschillen gevonden werden. De in zeven verschillende laboratoria gemeten metabolietconcentraties brachten verschillen tot een factor 3 naar voren, zelfs voor monsters die identiek waren behandeld. Interessant was dat metabolietverhoudingen beter vergelijkbaar bleken te zijn. Voor wat betreft de relatieve verschillen tussen de stammen of de groeicondities, gaven alle platforms vergelijkbare resultaten. Ook bepaling van enzymactiviteiten in twee verschillende laboratoria gaf vergelijkbare resultaten. Deze resultaten benadrukken het feit dat het verkrijgen van volledig reproduceerbare absolute metingen in de verschillende laboratoria nog steeds een uitdaging is, zelfs in afwezigheid van biologische of cultuurvariatie. Met behulp van de gegenereerde dataset was het ook mogelijk een integrale vergelijking van twee giststammen uit te voeren, waarin in het bijzonder de waargenomen verschillen in maximale groei en biomassaopbrengst werden onderzocht. We vonden aanwijzingen voor verhoogde expressie van zowel proteolytische als aminozuur biosynthese routes in de CEN.PK stam, waaruit blijkt dat een hogere turnover van eiwitten leidt tot de lagere biomassaopbrengst.

In **hoofdstuk 6** wordt een methode gepresenteerd waarmee, aan de hand van experimenteel verkregen gegevens de complexiteit van kinetische modellen kan worden beperkt. Aan de hand van een thermodynamische classificatie werden de individuele reacties van het metabole netwerk verdeeld in drie categorieën, gebaseerd op hun thermodynamische toestand en kinetische eigenschappen, namelijk: pseudo-evenwicht, nabij aan of ver van evenwicht. Deze aanpak wordt vervolgens gedemonstreerd met behulp van de technieken ontwikkeld en gevalideerd in de **hoofdstukken 2, 3 en 4**. Hiervoor werd een omvangrijke metabole dataset gegenereerd van hoge kwaliteit, voornamelijk bestaande uit intermediairen van het centrale metabolisme. De metingen werden verricht onder een reeks van 32 condities, waarbij de metabole fluxen maximaal een factor 60 van elkaar verschilden. Voor driekwart van de geanalyseerde reacties konden volledige van in-vivo data afgeleide wiskundige beschrijvingen worden verkregen, die direct konden worden toegepast in een kinetisch model. We vonden ook dat de *in vivo* kinetiek van de reacties geclassificeerd als bijna-evenwicht wordt gedomineerd door de thermodynamische drijvende kracht en dat deze eenvoudig beschreven kunnen worden als een lineaire functie van de verhouding tussen substraten en producten (Q), een format dat we "Q-linear kinetics" genoemd hebben. Deze vereenvoudiging bleek mogelijk te zijn omdat er een intrinsieke correlatie bestaat tussen de veranderingen in de reactant concentraties die zich voordoen *in vivo*. De gegenereerde dataset maakte het voor de eerste keer mogelijk om op een systematisch wijze Keq waarden voor biochemische reacties te verkrijgen onder in-vivo condities. Opmerkelijk hierbij is, dat een eerste vergelijking met gegevens van *E. coli* laat zien dat de aldus verkregen Keq waarden een betere referentie vormen voor nauwkeurige *in vivo* thermodynamische analyses, in vergelijking met de gebruikelijke in-vitro Keq data. Deze bevindingen tonen de mogelijkheid aan van datagedreven modelreductie door het aanpassen van de complexiteit van reactiesnelheids vergelijkingen aan de complexiteit van het daadwerkelijke kinetisch gedrag onder *in vivo* omstandigheden. De voorgestelde benadering biedt een kader voor de eerste fase van het opstellen van een kinetisch model waarbij slechts gebruik wordt gemaakt van beperkte a-priori informatie, als basis voor verdere kinetische modellering.

Table of contents

List of abbreviations

The following abbreviations are used throughout this thesis. Additionally, nucleotides and amino acids are referred to by their common 3/4-letter codes.

Metabolites:

2PG	2-phosphoglycerate	Gluc	glucose
3PG	3-phosphoglycerate	Isocit	isocitrate
6PG	6-phospho gluconate	M1P	mannose-1-phosphate
AcAld	acetaldehyde	M6P	mannose-6-phosphate
CIT	citrate	MAL	malate
DHAP	dihydroxyacetone phosphate	Mtl1P	mannitol-1-phosphate
E4P	erythrose-4-phosphate	OGL	oxoglutarate
EtOH	ethanol	PEP	phosphoenolpyruvate
F26bP	fructose-2,6-bisphosphate	Pi	phosphate
F6P	fructose-6-phosphate	PYR	pyruvate
FBP	fructose-1,6-bis-phosphate	R5P	ribose-5-phosphate
F2,6BP	fructose-2,6-bis-phosphate	Rbu5P	ribulose-5-phosphate
FUM	fumarate	S7P	sedoheptulose-7-phosphate
G1P	glucose-1-phosphate	SUC	succinate
G3P	glycerol-3-phosphate	T6P	trehalose-6-phosphate
G6P	glucose-6-phosphate	UDP-G	UDP-Glucose
GAP	glyceraldehyde-3-phosphate	X5P	xylulose-5-phosphate

Enzymes, or the reactions they catalyze:

ACO	aconitate hydratase (also known as aconitase)
ADH	alcohol dehydrogenase
ADK	adenylate kinase (also known as myokinase
APT	alanine transaminase
ENO	phosphopyruvate hydratase (also known as enolase
FBA	Fructose-bisphosphate aldolase
FMH	fumarate hydratase (also known as fumarase
G3PDH	glycerol-3-phosphate dehydrogenase
G6PDH	glucose-6-phosphate dehydrogenase
GPP	glycerol-1-phosphatase
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GPM	Phosphoglycerate mutase
HXK	hexokinase
HXT	hexose transporter
M1PDH	mannitol-1-phosphate 5-dehydrogenase
PDC	Pyruvate decarboxylase

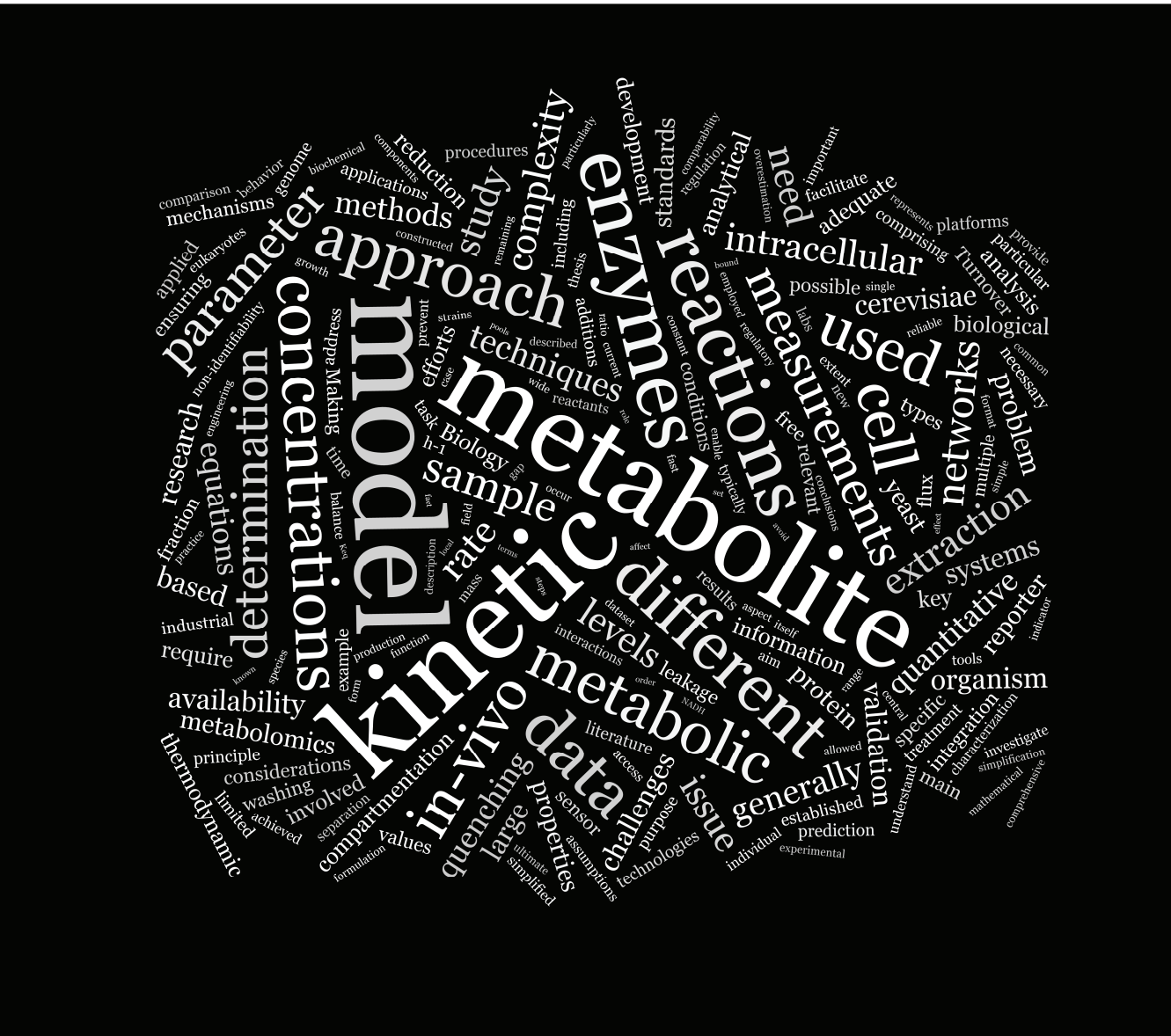
PFK	6-phosphofructokinase
PGI	Glucose-6-phosphate isomerase
PGK	phosphoglycerate kinase
PGM	Phosphoglucomutase
PMI	Mannose-6-phosphate isomerase
PMM	Phosphomannomutase
PYK	pyruvate kinase
RPE	Ribulose-phosphate 3-epimerase
RPI	Ribose-5-phosphate isomerase
TAL	transaldolase
TK1	transketolase (S7P-producing reaction)
TK2	transketolase (E4P-consuming reaction)
TPI	Triose-phosphate isomerase
TPP	trehalose-phosphatase
TPS	alpha,alpha-trehalose-phosphate synthase

Others:

a, b, p, q	substrates (a, b) and products (p, q) of a reaction, or their concentrations
CE	capillary electrophoresis
cyt	cytosolic
EX	extracellular
GC	gas chromatography
g_{DW}	gram of dry weight biomass
IC	intracellular
Keq	equilibrium constant
Ki	inhibition constant
Km	michaelis constant
LC	liquid chromatography
mit	mitochondrial
MS	mass spectrometry
q	biomass specific rate of production or consumption
Q	reaction quotient (also known as mass-action ratio)
QS	quenching solution
TOF	time-of-flight (mass spectrometry)
T, tot	total
v	reaction rate
Vmax	enzyme activity (the maximal velocity in the forward direction)
μ	specific growth rate

Chapter 1

General introduction



BACKGROUND

Metabolism can be understood as the combination of processes by which cells acquire and utilize energy to carry out their functions, from biosynthesis and cell division to communication and defense. This occurs via a network of literally hundreds of single biochemical reactions between metabolic intermediates, the behavior of which is controlled by multiple levels of interactions and regulatory mechanisms, from the genome to the individual enzymes (Figure 1.1). The study of metabolic reaction networks and the mechanisms regulating them not only extends our fundamental understanding of biology, it also expands the tools at our disposal for the prediction and rational engineering of cellular phenotypes.

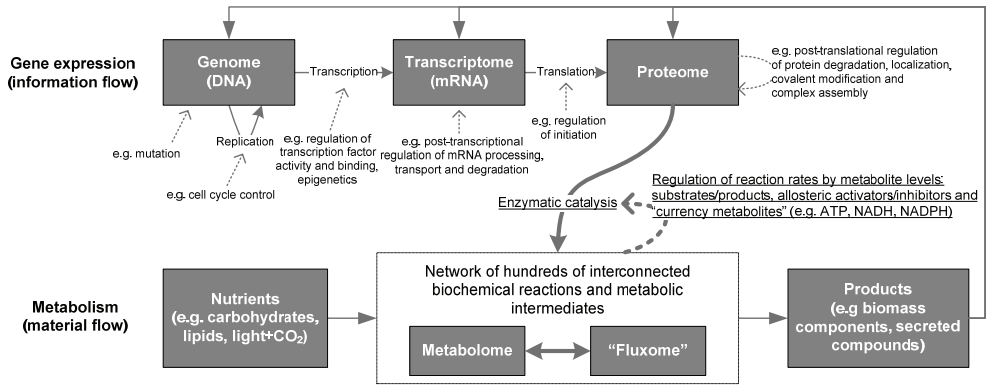


Figure 1.1: Simplified illustration of the organization of metabolism and the multiple levels of regulation, including examples of common types of regulation mechanisms (indicated by dashed lines). The work described in this thesis focuses on the level of the regulation of enzyme catalysis by the metabolites (underlined).

The realization of the multi-layered organization of regulatory mechanisms, of the sheer number of components involved and of the complex, non-linear nature of the interactions between them, is at the core of what can be considered the paradigm of Systems Biology: that an holistic approach is needed to understand and predict how the properties of biological systems emerge from the properties of their components. The overwhelming degree of complexity of metabolic systems makes intuitive prediction of their behavior practically impossible. This calls for an approach based on the formulation of large integrative mathematical models, which must be supported by the generation of comprehensive (multi-level) quantitative datasets. Ultimately, such integrated models will need to include descriptions of transcriptional regulatory networks, translational and post-translational machineries, protein interaction networks and enzyme kinetics. Making such an approach feasible will require tremendous advances at the technical, analytical and theoretical levels. It will also entail a (re-)shift in the way biology is studied, towards more quantitative methodologies.

The work described in this thesis focuses on the challenges surrounding the kinetic and thermodynamic analysis of metabolic reaction networks, with the ultimate goal of enabling the detailed network-wide mathematical modeling of enzyme kinetics *in vivo*.

Kinetic modeling of metabolic reaction networks

Within the integrative modeling approach envisioned in Systems Biology, a central and vital component is the modeling of the kinetics of enzymes in metabolic reaction networks. In fact, the construction of predictive quantitative kinetic models of metabolic pathways is an old ambition in Biochemistry which, despite repeated efforts, still remains largely unfulfilled.

A kinetic model normally consists of a set of mass balances around the metabolic intermediates (in the form of ordinary differential equations), comprising the mathematical description of the structure of the network (in the form of the stoichiometric matrix) and the kinetic properties of the individual enzymes (in the form of rate equations). This can be conveniently defined in matrix notation as:

$$\frac{dx}{dt} = S.v - \mu.x \approx S.v \qquad \text{(equation 1.1)}$$

The formalism is similar to Metabolic Flux Analysis except that kinetic models are by definition dynamic, so the steady-state assumption (dx/dt=0) need not apply.

The kinetic properties of the enzymes are typically expressed in the form of rate equations, which relate reaction rates with metabolite concentrations, via kinetic parameters. Rate equations can be derived for individual enzymes based on mechanistic considerations. For example, for one of the most common reaction types, a+b↔p+q via a ping-pong mechanism, the following rate equation can be deduced following standard procedures [39]:

$$v = V_{\max} \frac{\frac{ab}{K_{ia}K_{mb}}}{\frac{a}{K_{ia}} + \frac{K_{ma}.b}{K_{ia}K_{mb}} + \frac{p}{K_{ip}} + \frac{K_{mp}.q}{K_{ip}K_{mq}} + \frac{ab}{K_{ia}K_{mb}} + \frac{ap}{K_{ia}K_{ip}} + \frac{pq}{K_{ip}K_{mq}} + \frac{K_{ma}.bq}{K_{ia}K_{mb}K_{iq}}} \left(1 - \frac{\frac{pq}{ab}}{K_{eq}}\right) \qquad \text{(equation 1.2)}$$

This (relatively simple) rate equation involves one capacity parameter (Vmax), one thermodynamic parameter (Keq) and 7 affinity parameters (Km and Ki). For reactions involving more complex effects, such as cooperative binding or allosteric interactions, the mechanistic rate equations can be considerably more elaborate. Thus, a genome-scale kinetic model comprised of hundreds of non-linear rate equations might involve over 1000 kinetic parameters!

Assuming that the stoichiometry of the network is known, the key tasks in kinetic modeling are the formulation of the rate equations and their parameterization. The study of enzyme kinetics is an established research field with a long tradition so in principle it should be possible to obtain the necessary mechanistic information and parameter values from literature. In practice, kinetic information is scarce, incomplete or inexistent for most enzymes of most species (and tissues). Even when it is available, differences in strains or growth conditions limit their usefulness. In principle, enzymes from a single source can be extracted, purified and characterized. However, detailed kinetic (re-)characterization on a large scale would be extremely laborious. Most importantly, the possibility that the isolation itself affects enzyme properties, as well as the differences between the *in vitro* assay conditions and the intracellular environment, are sufficient to cast doubts on the validity of *in vitro*-derived parameters in

predicting *in vivo* behavior. This so-called "*in vivo* vs *in vitro* dilemma" is still a matter of debate. The main reason it hasn't been settled yet is the scarcity of the high-quality *in vivo* data needed to draw firm conclusions.

An alternative strategy is to characterize enzyme kinetics *in vivo* using whole cells. In principle, it should be possible to derive the meaningful *in vivo* kinetic parameter values for multiple enzymes simultaneously from measurements of fluxes, enzyme levels and metabolite concentrations. In practice, this is severely restricted by the limitations associated with generating the necessary *in vivo* data. Enzyme characterization *in vitro* relies on the ability to expose the enzyme to wide ranges of concentrations of reactants and effectors, independently and in a well-defined manner. In contrast, the ability to observe enzyme behavior *in vivo* is conditioned by the use of intact living cells, where enzymes are not operating in isolation and the concentrations of reactants and effectors cannot be manipulated at will. In addition, the ability to obtain the necessary measurements may also be limiting. The methodologies for determination of intracellular fluxes still evolving and the accurate, quantitative determination of intracellular metabolite concentrations and protein levels are still very demanding tasks.

That is, thus, the central dilemma of *in vivo* kinetic modeling: on the one hand, the size and complexity of the network implies the need for large models, comprised of non-linear rate equations and a multitude of parameters; on the other, the limited availability of or access to adequate *in vivo* data, which hampers the proper parameterization of such models. Attempting to estimate a large number of parameters from a limited dataset will generally result in parameter non-identifiability problems: infinite combinations of parameter values over wide ranges will be able to explain the data, so the values of individual parameters cannot be accurately determined. For most purposes that is not satisfactory, which is why preventing or resolving parameter non-identifiability issues must be a primary concern in kinetic modeling efforts.

All the key challenges to *in vivo* kinetic modeling, which will be enumerated below, can be understood in terms of the need to bridge this gap between model and data, by decreasing model complexity or by increasing data availability and quality.

Saccharomyces cerevisiae as model organism of industrial relevance

S. cerevisiae, also known as baker's yeast or budding yeast, has been used for millennia for brewing and baking. Today, it remains a particularly relevant subject of biological research thanks to its dual role as model organism and industrial production platform.

As the main model organism for unicellular eukaryotes, *S. cerevisiae* is also one of the most commonly used microbes in molecular and cell biology and, thus, one of the most characterized species. As an eukaryote, it shares the more complex, compartmentalized structure of plant and animal cells, but as a unicellular organism it is much simpler to study. It is also easy to cultivate, exhibits a high growth rate and has simple nutritional requirements. It is amenable to genetic manipulation, can be grown as haploid or diploid, and gene deletion collections have been established. It has a relatively small genome, which has been sequenced for at least two strains (with many others currently undergoing sequencing) and is among the

most extensively annotated. Finally, a variety of genomic and biochemical information and tools are available publicly, including databases such as SGD, MIPS/CYGD, YRC and YEASTRACT, which greatly facilitates data exchange and interpretation.

At par with *S. cerevisiae*'s usefulness in biological research is its utility for industrial applications. Its well-known traditional role in the brewing and bread industries is associated with its characteristic performance in the fast fermentation of sugars. More recently, *S. cerevisiae* is also playing an important role in Industrial Biotechnology as an attractive host for production of fuels (e.g. bio-ethanol) and other "cell-factory" applications. Its comparatively well-characterized biochemistry, the knowledge of the genome sequence and the availability of molecular biology tools make it better suited for genetic engineering than less conventional organisms. In addition, its relatively good tolerance to stresses, ease of cultivation, established use in large scales and Generally-Regarded-As-Safe status make it an attractive platform for large-scale production of biochemicals.

Thus, *S. cerevisiae* is a particularly interesting species for the study of metabolic reaction networks as well as the testing of Systems Biology approaches. The research is made easier by the knowledge and tools already available, while at the same time new discoveries may help answer long-standing questions, have an impact in our fundamental understanding of metabolic regulatory mechanisms, or facilitate future efforts towards the engineering of yeast, as well as other microbes, as efficient hosts for industrial production purposes.

FOUR KEY CHALLENGES FOR THE STUDY AND MODELING OF THE *IN VIVO* KINETICS OF METABOLIC NETWORKS

As mentioned, the key challenges to large-scale kinetic modeling can be understood in terms of the need to bridge the gap between model complexity and data availability. Reduction of model complexity to meet the availability of data will be considered as a single category, although it comprises several different approaches, each with its advantages and difficulties. More attention will be dedicated to discriminating the different types of technological challenges in generating comprehensive high-quality *in vivo* data and what is being done to tackle them.

Quantitative metabolomics

The kinetics and thermodynamics of biochemical reactions are, to a great extent, determined by the intracellular concentrations of the metabolites. Thus, any effort to understand and model metabolic reaction networks will eventually require absolute measurements of the concentrations of all relevant metabolites, including the reactants as well as effectors. That raises two major challenges: quantification and coverage.

The issue of coverage is mainly a matter of analytics. The metabolome of even simple microbes exceeds 1000 different compounds. Some can have very different chemical properties while others can be exceptionally similar (e.g. hexose phosphates, of which there are at least 20 mass isomers), and their intracellular concentrations can differ by at least 5 orders of

magnitude. Capturing the entire metabolome of an organism is thus an extremely demanding task, requiring simultaneously high resolution, sensitivity and throughput. Most current analytical platforms rely on MS detection coupled to GC, LC or CE separation. In fact, developments in these technologies have been the centre of attention and the main drivers behind the remarkable progress in the field of metabolomics over the last decade. Although for now claims in the literature are often exaggerated, kinetic modeling efforts are likely to drive as well as benefit from increased interest and progress in analytical metabolomics technologies.

As for quantification, accurate determination of intracellular metabolites requires not only quantitative analysis methods, which in the case of MS-based platforms implies the need for reliable isotope-based internal standards [28, 117, 209], but also adequate sampling and sample treatment techniques. An essential aspect in the determination of intracellular metabolites in biological samples is that their turnover is typically in the order of seconds (see Table 1.1). This fast turnover means that sub-second quenching of enzymatic activity is needed to preserve the "snapshot" of metabolism that a sample represents. It also implies that an emphasis on reliable sample treatment procedures is absolutely crucial to guarantee the quality of metabolite data.

Table 1.1: Turnover times of intracellular metabolites in *S. cerevisiae* growing under aerobic glucose-limited conditions, at two specific growth rates. Values are estimated from measured metabolite levels and calculated flux distributions (data from Chapters 6 and 4).

metabolite	Turnover time (s)		metabolite	Turnover time (s)		metabolite	Turnover time (s)	
	μ=0.1 h^{-1}	μ=0.38 h^{-1}		μ=0.1 h^{-1}	μ=0.38 h^{-1}		μ=0.1 h^{-1}	μ=0.38 h^{-1}
GAP	0.069	0.013	UDP	13	2.4	Met	95	42
2PG	1.7	0.028	G6P	17	2.5	Ile	210	44
PEP	6.7	0.037	6PG	6.3	2.6	Trp	330	60
E4P	0.2	0.057	Malate	12	2.8	T6P	64	210
ADP	0.36	0.059	Isocitrate	3.7	2.9	Phe	270	75
AMP	2.6	0.23	UMP	22	3	Glu	840	84
F6P	6.2	0.34	R5P	10	4	Val	550	100
ATP	1.3	0.36	X5P	10	5.1	Asp	590	120
3PG	15	0.38	NAD	5.2		Gly	210	130
DHAP	1.8	0.58	Pi	16	7.6	Cys	630	130
NADH	0.73		G3P	21	8.8	Tyr	570	140
M1P	6.5	0.77	Succinate	9.4	23	Asn	1000	160
Rbu5P	2.1	0.98	Citrate	42	9.5	Pro	1000	170
Pyruvate	1.1	1.3	S7P	110	11	Ser	230	190
Oxoglutarate	4.9	1.9	UDP-Gluc	68	12	Thr	300	230
G1P	6.1	2.1	UTP	33	13	Trehalose	35000	250
FBP	2.1	3.4	M6P	92	15	Gln	1300	280
Fumarate	3	2.2	Leu	150	37	Ala	2300	340

The state-of-the-art approach for sample treatment is depicted in Figure 1.2. The two crucial steps are the quenching/washing and the extraction. The purpose of rapid sampling an immediate quenching is to "freeze" enzyme activity (in reality it is not stopped but slowed down considerably), allowing sufficient time for the washing. Washing aims to remove metabolites present extracellularly as well as any contaminants in the medium that may interfere with analytical procedures. The extraction seeks to achieve complete cell disruption,

making all intracellular metabolites accessible for analysis, and inactivation of enzymes, to prevent unwanted changes in metabolite levels further on. Subsequent steps serve mainly to prepare the sample for analysis.

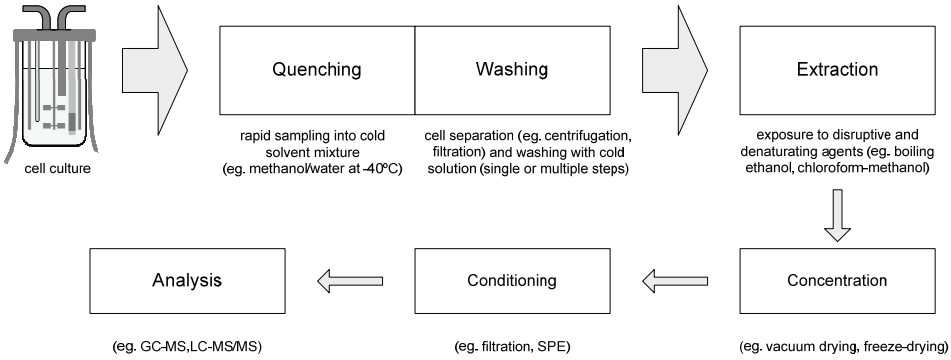


Figure 1.2: Overview of state-of-the-art metabolomics sample treatment for suspension cell cultures. Examples of common techniques are given below each unit operation. The steps until extraction should be carried out in succession, while afterwards samples can be stored.

One key challenge is to guarantee that quenching/washing causes neither underestimation nor overestimation of the intracellular metabolite pools. Overestimation can occur due to carry-over from the medium if the washing is not efficient enough. At typical cell densities, extracellular concentrations even in the order of mg/L can represent a large fraction compared to the intracellular amount, posing a non-negligible separation problem. Underestimation can occur if, upon exposure to solvents and/or low temperature, the cells release metabolites, a phenomenon known as "leakage" or "cold-shock" [188, 204]. The occurrence of leakage in eukaryotes is still disputed, but it is a well-accepted and yet unresolved problem in bacteria. Reliable quenching/washing thus requires protocol which ensures an adequate balance between washing efficiency and avoidance of leakage. The only way to draw firm conclusions on the occurrence of leakage or excessive carry-over is to measure metabolites in all sample fractions and establish mass balances. Surprisingly, this is seldom practiced, judging from the published literature.

Another key challenge is to ensure that the extraction is aggressive enough to ensure cell disruption and enzyme inactivation, but mild enough that metabolites are not entirely degraded. Furthermore, the resistance of different organisms and enzymes to varied extraction agents may differ so extraction techniques may need to be adjusted, or at least validated, for each organism. Traditionally, this validation involves the comparison of multiple extraction techniques as well as the determination of metabolite recoveries by spiking and/or standard additions. Presumably because this is cumbersome and care must be taken to avoid bias, several studies have relied instead on non-quantitative data or indirect evidence for method validation. This can ultimately lead to the adoption of inadequate procedures, as illustrated by the prominence that extraction by Freezing-Thawing in Methanol achieved in recent literature although, as we shall demonstrate, it does not satisfy the most basic criteria. Thus, insufficient method validation can have profound impact on the quality of metabolomics data and,

consequently, on the conclusions derived from such information. An important task is therefore the development of techniques and approaches for fast and reliable method validation and optimization.

Compartmentation and protein-binding

One of the most intractable challenges to generating meaningful *in vivo* kinetic data, on metabolite concentrations as well as protein levels, is the compartmentalization of eukaryotic cells (see Figure 1.3). Some cellular organelles (e.g. mitochondria, vacuoles) are surrounded by selective membranes and therefore should be regarded as separate compartments with specific pools of metabolites and proteins. In fact, differences in organelle composition are thought to be inherent to the specialization of metabolic function of the different organelles. Additionally, inter-compartment transport may be particularly relevant in determining certain system properties. To adequately understand and describe those functions, measurements at sub-compartment resolution will be needed.

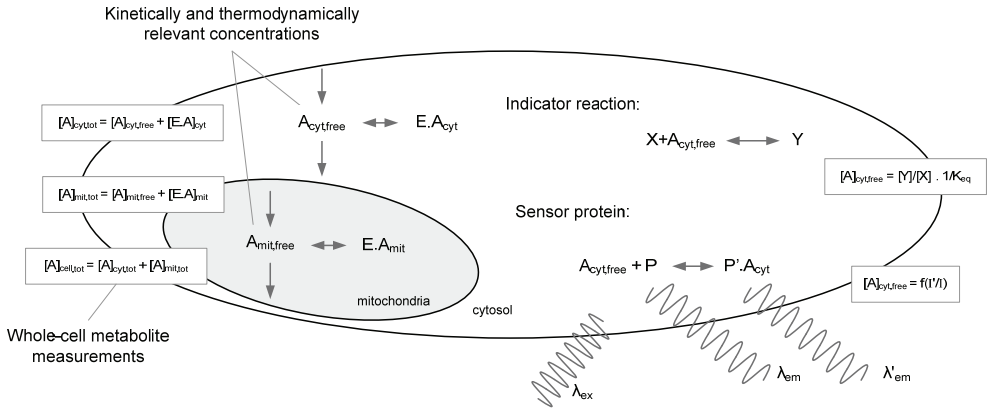


Figure 1.3: An illustration of the problems arising from cell compartmentation and protein-binding for the determination of relevant metabolite concentrations and simplified examples of two strategies to overcome them. The metabolite of interest (A) is present in both cytosol and mitochondria and tightly bound by enzymes (E), so its free concentrations in either compartment cannot be correctly assessed from whole-cell measurements. If a suitable indicator reaction can be found/introduced, the local free concentration of A can be determined from the measured concentrations of the remaining reactants (Y and X) and the equilibrium constant (Keq). Alternatively, if a suitable sensor protein is constructed, which emits light at different wavelengths (λ and λ') whether it is bound to A or not, the local free concentration of A can be determined from the difference in intensity (I) detected at each wavelength.

An issue somewhat related to compartmentation but that can affect metabolite measurements in both prokaryotes and eukaryotes is metabolite binding by proteins. Normally it is assumed that the fraction of each metabolite that is enzyme-bound is negligible. However, in the case of low concentration intermediates involved in many reactions catalyzed by very abundant enzymes (e.g. NADH, ADP, Mg^{2+}) the protein-bound fraction may be considerable. If so, the measured concentrations (free+bound) will largely overestimate the kinetically and thermodynamically meaningful free concentrations.

Yet, current methods for quantitative metabolomics are based on the complete disruption of the cells, thus yielding whole-cell average concentrations. So will quantitative proteomics

approaches that take the fast turnover of protein modification into account. On the other hand, common cell fractionation techniques based on homogenization and differential centrifugation are not compatible with the need to prevent enzyme activity. The only exception is perhaps a non-aqueous fractionation technique that has been applied in the study of plant cell compartmentation [53, 148]. If the same approach can be applied to other cell types, it could provide the answer to the compartmentation problem, though not to the protein-binding issue. An important task will be ensuring adequate validation and reproducibility, as well as streamlining the considerable work-load involved per sample for more high-throughput applications.

One alternative approach is to bypass direct measurements altogether by using genetically-encoded reporter reactions or sensor proteins. In principle, these can provide measurements of free concentrations at sub-cellular resolution. For example, pH-sensitive GFP-derived sensors have recently been employed to estimate the pH in different cell compartments of yeast [134]. Additionally, it is already possible to design fluorescence resonance energy transfer (FRET) sensors for specific targets via protein engineering, potentially enabling the determination of a wide range of metabolites [25, 46]. A different strategy is to use reporter reactions, allowing the determination of local metabolite ratios (e.g. NADH/NAD, NADPH/NADP) from the measured concentrations of the remaining reactants [182, 199]. This principle has been sometimes employed in mammalian liver cells but its application in other cell types is generally precluded by the lack of suitable reporters. A major constraint for both sensor and reporter based approaches is their poor scalability: a specific sensor/reporter must be separately designed, expressed and validated for every target molecule or ratio. Additional concerns include ensuring the specificity of the sensor/reporter and avoiding interference with the concentration of target compound itself as well as with the rest of metabolism.

Standardization and data integration

Ensuring that results can be reproduced between different labs is an essential aspect of scientific research and a particularly important concern in emerging fields, where techniques are still under development. The last few years have witnessed the establishment of several Minimal Information standards for specific types of experiments and measurements (e.g. MIAME, MIAPE, MIAMET), even though this is necessarily an iterative process as methodologies evolve and technologies mature. A particular difficulty in modern biological research is that the infrastructure and know-how required for high-throughput measurement technologies represents a sizable investment. There is therefore a tendency in most labs to specialize in a few (or single) technologies and to seek access to other types of measurement via inter-lab collaborations. Another aspect is the drive towards mass storage of experimental datasets in publicly accessible databases to facilitate access to information and data integration. This means that large-scale modeling efforts will often be based on experimental data collected from multiple sources, which is generally not straightforward.

Via Minimal Information standards it should be possible to address the issues of reporting and the creation of databases can facilitate data exchange. A different issue is ensuring that measurements generated by different labs are compatible, especially when slightly different

sample preparation procedures or different analytical platforms are employed. If differences in results are observed, finding the sources and solving the problem may require substantial effort. The few studies reporting inter-laboratory or inter-platform comparison, involving semi-quantitative or profiling techniques, offer somewhat optimistic prospects [78, 88, 137, 159, 198]. Since the action of enzymes is determined by the concentrations of molecules rather than ratios or fold-changes, a crucial task will be achieving comparability at the level of absolute quantification.

Reduction of model complexity

Model reduction is a field by itself. It has its roots in statistics and control theory and is most commonly applied to models of electrical and mechanical systems. The most basic motivation for model reduction is the need for simple models that capture the main features of complex dynamic systems but are easier to compute, thus facilitating simulation and control applications. With biological systems, model reduction has typically been applied *ad hoc* to adjust model complexity in view of the shortcomings in terms of availability of *in vivo* data. It can also be used to extract the components and interactions most relevant to the system's properties with the purpose of facilitating interpretation [84] or to systematically address parameter non-identifiability issues (e.g. [124, 126]).

Among different methods of reducing the complexity of kinetic models, three main approaches can be distinguished: a) simplifying assumptions, such as lumping sets of reactions or selectively setting constant values to rates, metabolite levels, or metabolite sums (e.g. conserved moieties); b) formulating rate equations using approximative kinetic formats, such as Generalized Mass-Action, non-competitive Michaelis-Menten, Power-Law and Linlog [27, 74]; and c) applying quasi-steady-state and/or rapid-equilibrium approximations based on thermodynamic or time-scale considerations [75].

A key issue in model reduction is achieving an adequate balance between detail and simplification: insufficient reduction may fail to resolve parameter non-identifiability problems, while too much simplification can lead to loss of system properties and predictive capacity**.** A closely related issue is the validity and relevance of the simplifications made on the basis of assumption or theoretical considerations. For example, the most commonly used simplifying assumptions (e.g. pseudo-equilibrium, constant conserved moieties, constant pH, lumping of large sections of metabolism) are typically employed by necessity rather than based on experimental evidence, and it is usually unclear to what extent they may influence the outcomes of the model. On several occasions such assumptions are later found to be incorrect when the experimental observations do become available (e.g. [95, 96]). Regarding the use of approximative kinetic formats, while using a unified kinetic format has computational advantages, a "one-size-fits-all" approach may result in simultaneous over-complication of the description of certain reactions and over-simplification of others. Hybrid approaches (e.g. [27]) may be more successful at exploiting the potential for simplification afforded by the kinetics of each reaction.

AIM AND OUTLINE OF THE THESIS

The aim of this research project was to contribute to the construction of an *in vivo*-validated kinetic model of central metabolism in *S. cerevisiae*. The project itself was part of larger collaboration initiative between six research groups from three Dutch universities, entitled "*Vertical Genomics: from gene expression to function, … and back*", aiming to develop and test integrated analysis approaches comprising the multiple levels of regulation, from the genome to fluxes, with yeast glycolysis as object of study.

Each chapter of this thesis tackles a particular challenge among those enumerated above. A strong emphasis was placed on data-driven approaches, rooted on the generation of high-quality state-of-the-art quantitative data. Although the ultimate objective has not yet been fully achieved, it can be truthfully said that every one of the chapters in this thesis succeeds at solving an important question or problem. Each of them therefore makes a significant contribution to the development of techniques that will eventually enable the network-wide *in vivo* kinetic modeling of *S. cerevisiae*, as well as other biological systems.

The issue of quenching yeast cells is analyzed in detail in **Chapter 2**. In particular, the question of whether leakage of intracellular metabolites to the quenching solution occurs during conventional cold methanol quenching, and to what extent, is still debated in the literature. To investigate this issue in a strictly quantitative manner we established a comprehensive mass balance approach based on the determination of metabolite levels in all sample fractions (i.e. whole-broth, filtrate, quenched cell pellets and respective quenching solutions). Using this approach we were able to adequately quantify the extent of leakage and investigate the effect of varying the quenching conditions, with the purpose of developing an improved method that entirely prevented leakage. We were also able to investigate whether separation of the extracellular metabolites in the medium is necessary to prevent overestimation of intracellular pools.

In **Chapter 3** we deal with the aspect of metabolite extraction. In particular, previous studies reporting comparisons of extraction techniques are utterly contradictory regarding the performance (and even the adequacy) of each method. Furthermore, in most cases it is unclear to what extent the criteria used in those studies may be affecting their conclusions. To circumvent the limitation of traditional validation procedures (spiking and standards additions) and address the problem systematically we introduced a new approach for the determination of metabolite recoveries, based on differential additions of ^{13}C-labeled internal standards at different stages of sample treatment. This allowed us to evaluate the five main methods for metabolite extraction in a quantitative and unbiased manner. We were also able to provide new insight into whether a combination of extraction methods in parallel is necessary for the comprehensive analysis of different metabolite classes, a matter which is still debated in the literature.

The issue of compartmentation is addressed in **Chapter 4**, which describes the successful application of the indicator reaction principle to determine the free NAD/NADH ratio in the cytosol of *S. cerevisiae*. To circumvent the absence of adequate indicators in yeast, it was necessary to introduce a heterologous enzyme chosen specifically for that purpose. We were

able to investigate the impact of the new measurements on the thermodynamic analysis of key reactions, such as GAPDH. It was also possible to apply the new system to study the effect of sudden changes in the external supply of electron donors and acceptors on the redox state of the cytosolic NAD/NADH couple.

The inter-laboratory experiment described in **Chapter 5** addresses the issue of standardization and data integration. Carried out within the framework of the EU-funded Yeast Systems Biology Network, one of its main goals was to determine whether the analytical pipelines currently used at several European yeast research labs, which in many cases have been developed with particular research interests in mind, provide comparable results. Particular attention was given to transcriptomics, for which there are a several analytical methods in use, and metabolomics, for which both sample treatment protocols and analytical platforms vary substantially between labs. The comparison between two strains (CEN.PK and a newly constructed YSBN strain) and two growth conditions (chemostat and batch) provided a biological context for the comparison of results.

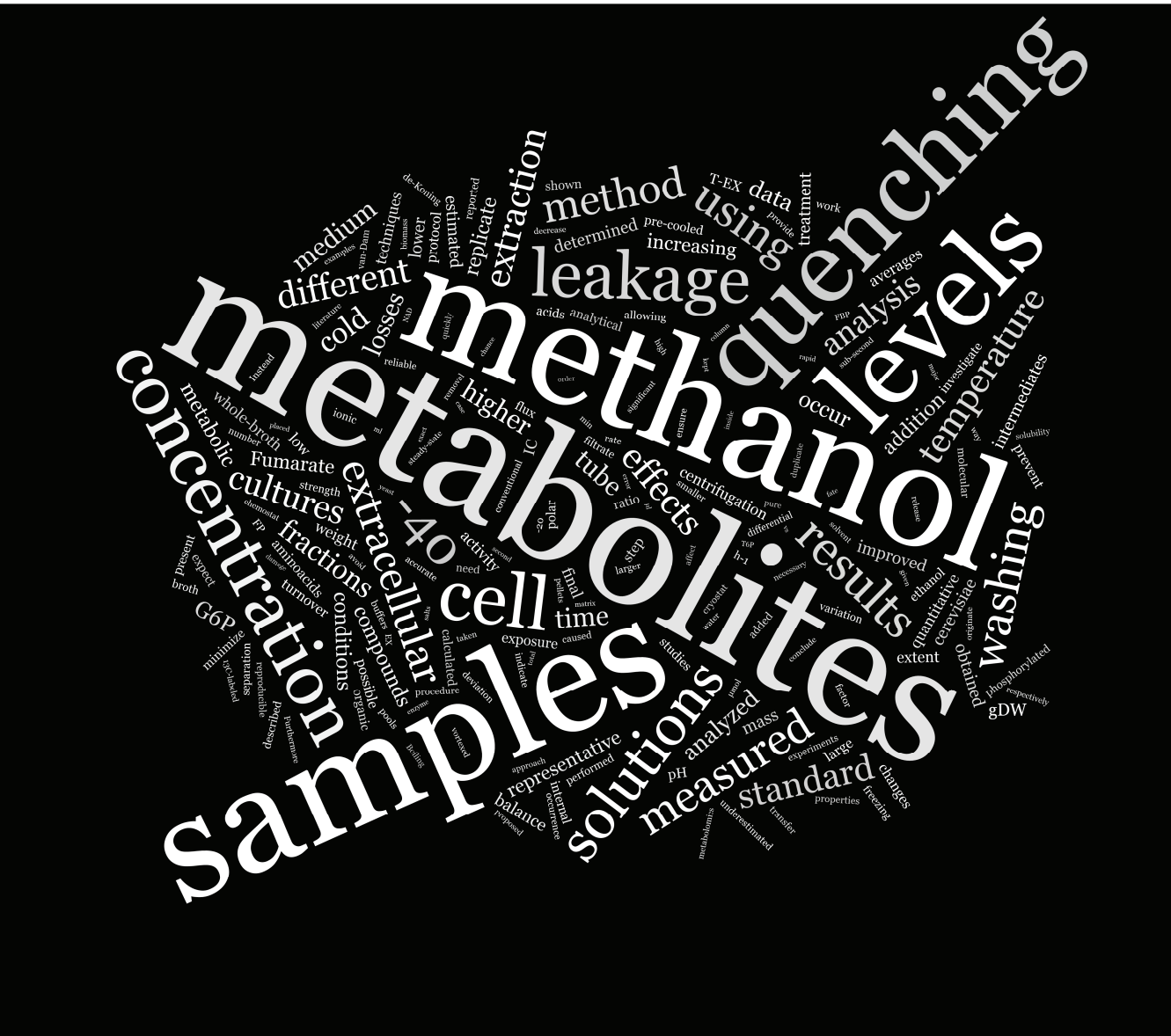
Making use to the tools and techniques developed in **Chapters 2, 3 and 4**, we addressed the problem of kinetic parameter non-identifiability and model reduction, from a data-driven perspective, in **Chapter 6**. We propose an approach that uses *in vivo* data to classify individual reactions, according to their thermodynamic state and kinetic behavior, into three categories: pseudo-, near- and far-from-equilibrium. The aim is to enable the degree of complexity in which each reaction is modeled to be precisely tailored to the complexity of its kinetics *in vivo*, thus closing the gap between model and data. To demonstrate the approach we used chemostat cultures to generate a large metabolite and flux dataset, comprising most of central metabolism, over a wide range of flux conditions. This allowed us to classify a total of 27 reactions. It was also possible to fully parameterize simplified mathematical descriptions for all reactions in the first two categories, which represented 3/4 of the reactions analyzed. Those parameters included apparent *in vivo* Keq values, which we investigated as potential references for *in vivo* thermodynamic analyses across different species. This study also led us to propose a convenient kinetic format, dubbed "Q-linear kinetics", to describe the reactions classified as near-equilibrium. The approach constitutes a useful framework for the initial stages of model formulation with minimal need for prior kinetic information on the enzymes.

Finally, some prospects for the continued efforts towards *in vivo* kinetic modeling of metabolic reaction networks are discussed in **Chapter 7**.

Chapter 2

The true contents of the cell

An improved, leakage-free quenching method for yeast metabolomics



ABSTRACT*

Accurate determination of intracellular metabolite levels requires reliable, reproducible techniques for sampling and sample treatment. Quenching in 60% (v/v) methanol at -40°C is currently the standard method for sub-second arrest of metabolic activity in microbial metabolomics but there have been contradictory reports in the literature on whether leakage of metabolites from the cells occurs. We have re-evaluated this method in *S. cerevisiae* using a comprehensive, strictly quantitative approach.

By determining the levels of a large range of metabolites in different sample fractions and establishing mass balances we could trace their fate during the quenching procedure and confirm that leakage of metabolites from yeast cells does occur during conventional cold methanol quenching, to such an extent that the levels of most metabolites have been previously underestimated by at least 2-fold. In addition, we found that the extent of leakage depends on the time of exposure, the temperature and the properties of the methanol solutions. Using the mass balance approach we could study the effect of different quenching conditions and demonstrate that leakage can be entirely prevented by quenching in pure methanol at ≤-40°C, which we propose as a new improved method. Making use of improved data on intracellular metabolite levels we also re-evaluated the need of sub-second quenching of metabolic activity and of removing the extracellular medium.

Our findings have serious implications for quantitative metabolomics-based fields such as non-stationary ^{13}C flux analysis, *in vivo* kinetic modeling and thermodynamic network analysis.

* Published as: Leakage-free rapid quenching technique for yeast metabolomics, in Metabolomics (2008) 4, 3: 226-239

INTRODUCTION

The accurate measurement of physiological levels of intracellular metabolites is of prime interest in the study of metabolic reaction networks and their regulation *in vivo*. Owing mainly to developments in analytical tools, in particular in MS-based techniques, the field of metabolite analysis is undergoing fast expansion. Metabolomics, the systematic analysis of large numbers of low molecular weight compounds from a biological system, is in the process of establishing itself as a global analysis method complementary to transcriptomics and proteomics and is expected to provide major contributions to areas such as functional genomics, toxicology and nutrigenomics. However, whether for quantitative or qualitative purposes, the quality and reliability of metabolomics data will invariably depend on the sampling and sample treatment techniques employed, which usually receive relatively little attention. Strikingly, there is no consensus in the literature on the effectiveness or even adequacy of the available techniques for sampling, quenching and extraction of intracellular metabolites from microbial cultures[16, 35, 45, 45, 67, 70, 72, 106, 109, 114, 128, 157, 187, 188, 204].

Many intermediates in metabolic reaction networks have turnover times in the order of seconds, due to the relatively high conversion rates and low metabolite concentrations found *in vivo*. The need to rapidly quench metabolic activity upon sampling to avoid unwanted changes in intracellular metabolite levels was already recognized and documented by biochemists many decades ago [54, 182, 199, 206]. In the case of cell suspension cultures rapid sampling was initially accomplished by sampling broth directly into a cold perchloric acid (PCA) extraction solution followed by a series of freezing-thawing cycles, thus achieving quenching, release of intracellular metabolites and inactivation of enzymes [40, 73, 195]. However, direct extraction has two major disadvantages: the low concentrations of metabolites in the samples, owing to low biomass densities; and the risk of overestimating intracellular pools, because the metabolites present in the extracellular medium are not removed. The first significant attempt to combine quenching with cell separation involved fast filtration and washing of the cells with 50% (v/v) methanol at -40°C, followed by the extraction step [153]. Maintaining a very low temperature minimized metabolic activity during filtration while allowing removal of the extracellular medium. This technique was later improved by De Koning and Van Dam, who proposed sampling the broth directly into 60% (v/v) methanol at -40°C and separating the cells by centrifugation[45]. That was a decisive improvement because it allowed sub-second arrest enzymatic activity. This technique has remained mostly unchanged and is still the most widespread method for rapid sampling of microbial cultures [114, 128].

Perhaps the most critical assumption in the cold methanol quenching method is that intracellular metabolites will remain inside the cells during quenching and centrifugation. If metabolites were to leak from the cells into the methanol solutions, which are discarded, the intracellular levels might be severely underestimated. The original work of De Koning and Van Dam and some of the subsequent literature concluded that metabolite leakage did not occur in yeast [45, 67], fungi [70, 152] or bacteria [121]. However, this assumption is increasingly being questioned after some recent studies, making use of more sensitive MS-based analytical methods, reported

the occurrence of extensive losses of intracellular metabolites during cold methanol quenching in yeast [188] as well as bacteria [16, 204]. Unfortunately, this issue has not been thoroughly and systematically addressed yet. In addition, if leakage does occur, there is no reliable, validated alternative method, despite some recent efforts in finding substitute quenching solutions[187].

The purpose of this work was to determine quantitatively whether leakage of intracellular metabolites occurs in *S. cerevisiae* during cold methanol quenching and, if so, find a way to prevent it. In contrast with previous method evaluation attempts, we proposed to use only strictly quantitative data from samples obtained in standardized, reproducible culture conditions through well-defined rapid sampling and sample treatment procedures. In addition, a wide range of metabolites was analyzed, to cover different classes of compounds (phosphorylated intermediates, organic acids, aminoacids) and different molecular properties (e.g. molecular weight, polarity). Furthermore, after initial indications of leakage we adopted an even broader approach by measuring metabolite levels in all possible sample fractions and establishing mass balances to be able to trace the fate of the metabolites during cold methanol treatment. Only such a comprehensive approach made it possible to positively conclude on the occurrence of leakage under different conditions and find a way to prevent it. We hope the outcomes of this work will serve those in the metabolomics community by providing an improved quenching method for intracellular metabolite analysis in *S. cerevisiae* and a standard for method validation in other microorganisms.

EXPERIMENTAL PROCEDURES

Solvents and chemicals - HPLC-grade methanol, ethanol and chloroform were supplied by Baker (The Netherlands). Analytical grade standards were supplied by Sigma.

Strain and cultivation conditions - The *Saccharomyces cerevisiae* strain used in this study was CEN.PK 113-7D (MATa) [176]**.** The cells were grown in aerobic carbon-limited chemostat cultures in a 7 L fermentor (Applikon, The Netherlands) with a working volume of 4 L. Unless stated otherwise, defined mineral medium [185] with 7.5 g/L glucose was used, which supported a steady-state biomass concentration of 3.7g_{DW}/L, the dilution rate was 0.1 h^{-1} and the aeration rate was 0.5 vvm (120 L/h). Dissolved oxygen tension (DOT) was measured in-situ with an oxygen probe (Mettler-Toledo, Switzerland) and O_2 and CO_2 concentrations in the off-gas were measured at-line using a combined paramagnetic/infrared analyzer (NGA 2000, Rosemount, USA). The pH was controlled at 5.0 with 4 M KOH and the temperature was set at 30°C. The overpressure in the vessel was kept at 0.3 bar and the stirrer speed was 600 rpm, ensuring that the DO was always above 80%. All experiments were carried out with steady-state cultures, that is, after 5 residence times of glucose-limited growth with constant DOT and off-gas readings.

Samples for intracellular metabolites (IC) - Samples were taken using a specialized rapid-sampling setup [100]. Unless stated otherwise, approximately 1 g (±0.05) of broth was withdrawn and injected (≤0.8 s) into a tube containing 5 ml 60% aqueous methanol (v/v) solution pre-cooled to -40°C, the contents of the tube were quickly mixed by vortexing (≈1 s) and the tube

was placed back in the cryostat at -40°C (Lauda, Germany). This way, a set of replicate samples can be taken quickly without noticeably disturbing the steady-state of the culture. During all subsequent steps the temperature of the tubes was maintained as close to -40°C as possible. The exact sample weights were determined by weighing each tube before and after sampling. The tubes were centrifuged using a rotor pre-cooled to -40°C at 4000 G for 5 min in a centrifuge cooled to -20°C. Unless stated otherwise, the biomass pellets were washed in 5 ml 60% methanol (v/v) solution pre-cooled to -40°C and centrifuged again as before. After decanting, U-^{13}C-labeled cell extract was added to the cell pellets as internal standard [117, 209].

Metabolite extraction - Extraction of intracellular metabolites was performed using the boiling ethanol method, adapted from [67] as described in [100]. Briefly, each tube was taken from the cryostat at -40°C and 5 ml 75% (v/v) boiling ethanol was added. Each tube was immediately vortexed and placed in a water bath at 95°C. After 3 min each tube was placed back in the cryostat.

Sample concentration - All ethanol extracts were evaporated under vacuum for 110 min, as described in [117]. Dried residues were resuspended in 500 µl demineralized water and centrifuged at 15000 G for 5 min at 4°C. The supernatants were stored at -80°C until analysis.

Samples for quenching solution (QS) and washing solution (WS) metabolites - Instead of discarding the methanol supernatants after quenching or washing, they were collected in pre-cooled tubes at -40°C. Each methanol solution was thoroughly vortexed and 300-500 µl was transferred to an empty pre-cooled tube. The exact sample weights were determined by weighing all tubes before and after transfer and keeping record of the corresponding intracellular samples. U-^{13}C-labeled cell extract was added as internal standard. Boiling ethanol extraction was performed as above to minimize the chance of sample matrix effects and ensure inactivation of any enzyme activity.

Samples for whole-broth (T) metabolites - Sampling was done as for intracellular metabolites but the quenched cell suspension was not centrifuged. Instead, it was thoroughly vortexed and 300-500 µl was transferred to an empty tube pre-cooled to -40°C. The exact sample weights were determined by weighing all tubes before and after sampling and transfer. U-^{13}C-labeled cell extract was added as internal standard. Boiling ethanol extraction was performed as above to minimize the chance of sample matrix effects and ensure complete cell disruption and inactivation of enzyme activity.

Samples for extracellular (EX) metabolites - Broth was quickly sampled by over-pressure into a syringe containing an amount of cooled steel beads designed to bring the temperature down to 0°C [115]. The broth was then quickly filtered through a 0,45 µm cartridge filter directly into a tube containing 5 ml 60% methanol (v/v) at -40°C. The resulting methanol filtrate solution was thoroughly vortexed and 300-500 µl was transferred to an empty tube pre-cooled to -40°C. The exact sample weights were determined by weighing all tubes before and after sampling and transfer. U-^{13}C-labeled cell extract was added as internal standard. Boiling ethanol extraction was performed as above to minimize the chance of sample matrix effects and ensure inactivation of any enzyme activity.

Metabolite analysis - The concentrations of the metabolic intermediates G6P, F6P, FBP, PEP, pyruvate, T6P, 6PG, G1P, M6P, citrate, oxoglutarate, succinate, fumarate, and malate, as well as the combined pool 2PG+3PG, were determined by ESI-LC-MS/MS [172]. The concentrations of several aminoacids were determined by GC-MS using the EZ:Faast kit for free aminoacid analysis from Phenomenex (Torrance, CA, USA). Quantification of the metabolites was based on the use of U-^{13}C-labeled cell extract as internal standard [117, 209].

RESULTS AND DISCUSSION

Effect of prolonged exposure

Our initial experiments aimed at reproducing the results of De Koning and Van Dam, who ruled out the occurrence of leakage because measured metabolite levels did not decrease significantly after 30min of extra exposure to the quenching methanol solution[45]. We tested this by placing samples back in the cryostat after re-suspension in the washing methanol solution and leaving them at -40°C for periods of +0, +30, +60 or +90min prior to the second centrifugation step. Unlike the majority of the experiments reported here, this was done with samples from cultures at D=0.05 h^{-1} and a biomass concentration of 14.5 g_{DW}/l [96]. If leakage did not occur, we would expect to find the same intracellular levels regardless of the time of exposure. The results are shown in Figure 2.1. G6P (MW=260) and Fumarate (MW=116) are given as representative examples of the results found for phosphorylated intermediates (larger, more polar) and organic acids (smaller, less polar), respectively.

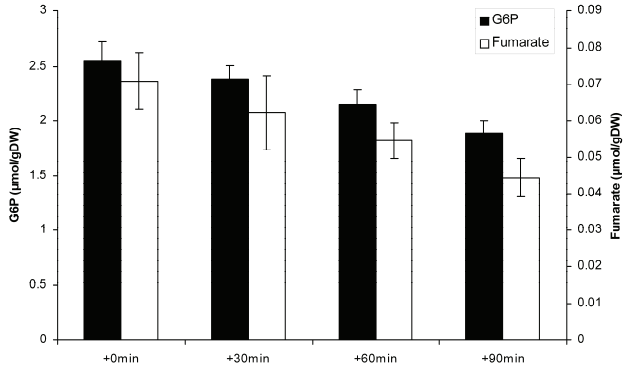


Figure 2.1: Effect of prolonged exposure to methanol washing solution (60% v/v) on the measured intracellular levels, exemplified for G6P and fumarate. Prolonged exposure was achieved by leaving the samples at -40°C before the second centrifugation step. Samples were from a culture at D=0.05 h^{-1} and 14.5 g_{DW}/l [96]. Data are averages ± standard deviation of 4 replicate samples, each analyzed in duplicate

The results showed that the longer the cells remained in contact with the washing solution, the lower the measured intracellular levels, suggesting time-dependent release of the intracellular metabolites. Interestingly, the rate of decrease was somewhat higher for smaller compounds, which might be released more easily, than for larger compounds. All samples were replicates, taken, treated and analyzed as one sample-set. Sample matrix effects in the analysis can be

discarded since U-^{13}C-labelled extract was used as internal standard. Then, differences can only originate from the period of exposure to the methanol washing solution. These results are not necessarily in contradiction with the work of De Koning and Van Dam, since the measured losses represented not more than 30% per hour (so <15% in 30min), which were probably not observable with the enzyme-based analytical techniques available at the time. Nevertheless, they do indicate that leakage is occurring and that the contact time with the methanol solution should be kept to a minimum. That is broadly in agreement with the findings of Villas-Boas et al. [188]. What cannot be concluded from these results is whether significant losses occur even without prolonged exposure, since it cannot be assumed that the rate of loss is the same throughout the entire procedure.

Effect of buffers and ionic strength

Although the original method proposed by De Koning and Van Dam described the use of non-buffered methanol solutions for quenching and washing, much of the later literature describes the use of methanol solutions with some sort of additive, usually a buffer [67, 152]. Presumably, this would prevent or minimize cell damage caused by changes in pH or osmotic shock. However, to our knowledge no quantitative data has ever been presented to back up this claim. We therefore tested the effect of adding buffers (HEPES at pH 5 or Tricine at pH 6) or salts (NH_4HCO_3, pH 8), at two different concentrations (10 or 100 mM), by quenching and washing replicate samples in each type of solution and comparing the metabolite levels measured in the resulting intracellular extracts. If changes in pH or low ionic strength caused cell damage and losses of metabolites from the cells, we would expect to see higher levels of intracellular metabolites in samples treated with buffers or salts. The results are shown in Figure 2.2. Again, G6P and Fumarate are given as representative examples of the results found for phosphorylated intermediates and organic acids, respectively.

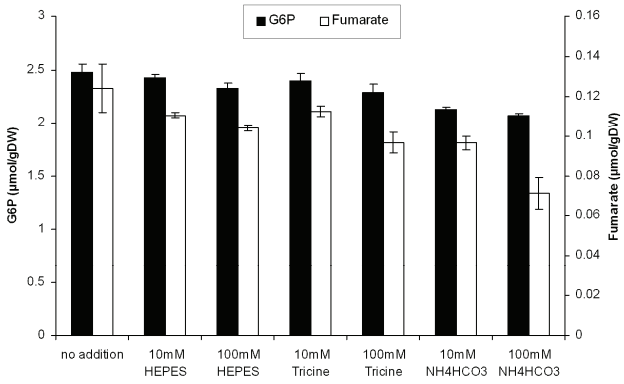


Figure 2.2: Effect of additions (HEPES at pH 5, Tricine at pH 6 or NH_4HCO_3 at pH 8, to 10 or 100mM) to the methanol quenching/washing solutions (60% v/v) on the measured intracellular levels, exemplified for G6P and fumarate. Samples were from a culture at D=0.1 h-1 and 3.7 gDW/l. Data are averages ± standard deviation of 2 replicate samples, each analyzed in duplicate

The results show that there is no significant benefit in buffering or increasing the ionic strength of the methanol solution. On the contrary, for most metabolites this resulted in slightly lower intracellular amounts. In addition, this effect was more pronounced for smaller

compounds, which might be released more easily, than for larger compounds. All samples were replicates and U-^{13}C-labelled extract was used as internal standard, so the differences must originate from the quenching and washing steps. These results suggest that losses can occur during quenching/washing, since changing the properties of the methanol solutions can affect the levels found in the cell pellets. Furthermore, they showed that if losses are occurring, they cannot be prevented by adding buffers or salts to the methanol solutions. On the contrary, such additions may even enhance the extent of the losses.

Effect of methanol concentration

Methanol is widely regarded as a toxic compound that can have detrimental effects on cell membrane integrity. A methanol concentration of 60% (v/v), as proposed by De Koning and Van Dam, is just enough to keep the cell suspension after sampling (50% v/v final, FP≈-42°C [44]) from freezing at -40°C. This seems to reflect the perceived need to keep the methanol concentration as low as possible, for fear of causing damage to the cells. We are not aware of subsequent literature where this has been changed. However, to our knowledge no quantitative data has ever been presented to demonstrate the adverse effects of methanol in the context of quenching and intracellular metabolite analysis. We therefore evaluated the effect of the methanol concentration by quenching and washing replicate samples in 50, 60, 70 and 80% methanol and comparing the metabolite levels measured in the resulting intracellular extracts. To avoid freezing of the samples quenched with 50% (v/v) solutions (42% v/v final, FP≈-32°C [44]), in this experiment the cryostat temperature was -30°C instead of -40°C. If increasing methanol concentrations caused cell membrane damage and this allowed release of intracellular metabolites, we would expect to find higher metabolite levels in samples treated with lower methanol concentrations. The results are shown in Figure 2.3**.** Once again, G6P and Fumarate are given as representative examples of the results found for phosphorylated intermediates and organic acids, respectively.

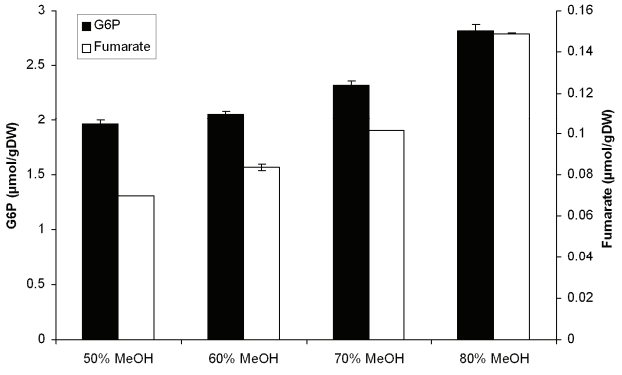


Figure 2.3: Effect of the methanol (MeOH) concentration of the quenching/washing solutions on the measured intracellular levels, exemplified for G6P and fumarate. To avoid freezing of the samples quenched with 50% methanol the cryostat temperature was -30°C instead of -40°C. Samples were from a culture at D=0.1 h-1 and 3.7 gDW/l. Data are averages ± standard deviation of 2 replicate samples, each analyzed in duplicate

Contrary to our expectations, in samples obtained by quenching and washing with higher methanol concentrations the intracellular metabolites levels were higher. In addition, the

differences followed a clear, smooth profile as function of increasing methanol concentration. All samples were replicates and U-^{13}C-labelled extract was used as internal standard, so the differences must originate from the quenching and washing steps. Like before, the fact that changing the properties of the methanol solutions could affect the obtained intracellular levels of metabolites suggested that losses can occur during quenching/washing. More importantly, the results unexpectedly show that under these conditions higher concentrations of methanol may reduce those losses, rather than increase them.

Full mass balance and the importance of methanol concentration and temperature

Although useful information can be obtained by analyzing only the intracellular samples, as presented above, these provide merely indications. To draw definitive conclusions on the occurrence and the extent of leakage it is necessary to analyze metabolite levels in all the other sample fractions and perform mass balances. Only then can the fate of the metabolites during sample treatment be adequately evaluated. To achieve this, we have analyzed metabolite amounts in 6 different fractions (see Figure 2.4), representing washed and non-washed intracellular samples (IC), the respective quenching and washing methanol solutions (QS and WS), the medium filtrate (EX) and the whole-broth (T).

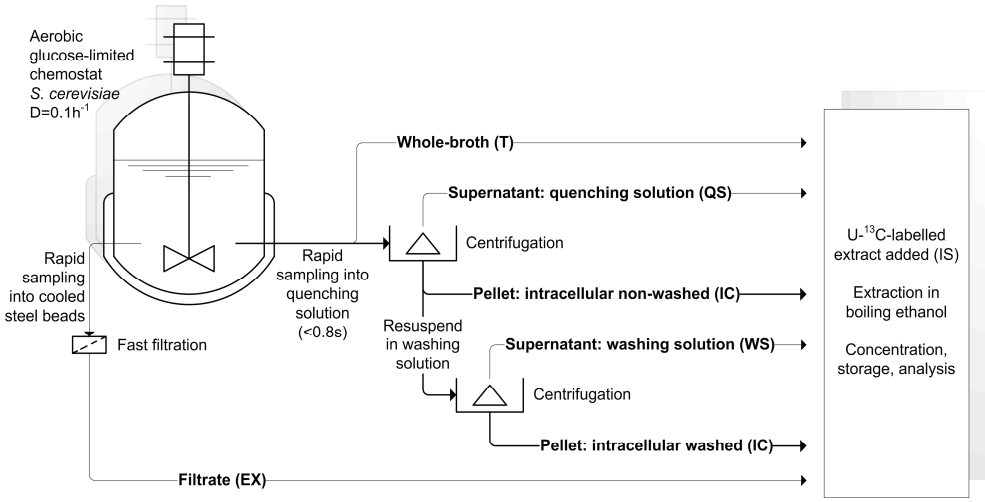


Figure 2.4: Sample fractions where metabolite concentrations were measured to investigate the fate of the metabolites during sample treatment (see Methods for details). The temperature and methanol concentration of the quenching/washing solutions was varied according to

The actual metabolite levels in the cells were estimated by the so-called "differential method" by subtracting the amount in the extracellular medium from the whole-broth total (T-EX). Furthermore, we analyzed all fractions for 8 variations of the cold methanol quenching

protocol, as described in Table 2.1, to investigate the effects of methanol concentration and temperature.

These samples were analyzed not only for phosphorylated intermediates and organic acids but also for aminoacids, providing an even wider range of compounds. All metabolite levels were expressed in μmol/gDW, to establish the mass balances. The mass balance results are shown in Figure 2.5. G6P and Fumarate are given as representative examples of the results found for phosphorylated intermediates and organic acids and Glutamate (MW=147) as a representative of mid-molecular weight aminoacids. Average mass balance closure over all measured metabolites (calculated as (IC+QS)/T) was 105% (±36%), which is quite satisfactory considering the analytical challenge involved. The cell levels calculated by the differential method (T-EX) can be used as a benchmark, with which the levels measured in the methanol-quenched intracellular samples (IC) can be compared. Detailed data on these values, for each measured metabolite, is provided in Table 2.2.

Table 2.1: Variations in sample treatment protocol tested to investigate the effects of temperature and methanol concentration on metabolite leakage (see Figure 2.2).

Protocol variation	Temperature	Conc. of methanol solutions (v/v)	Sample/quenching solution ratio	Conc. of methanol after sampling (v/v)	Washing step
A	-20°C	40%	1:5	33%	no
B	-20°C	60%	1:5	50%	no
C	-40°C	60%	1:5	50%	yes
D	-40°C	60%	1:5	50%	no
E	-40°C	80%	1:5	67%	no
F	-40°C	100%	1:5	83%	no
G	-40°C	100%	1:10	91%	no
H	-78°C	100%	1:10	91%	no

The IC/(T-EX) ratio, which is a measure of how close each protocol variation comes to delivering to "true" intracellular levels, is represented in Figure 2.6. The results show that leakage of intracellular metabolites does occur in the standard procedure of quenching in 60% methanol at -40°C (Figure 2.5), regardless of whether a washing step is carried out (C) or not (D). The measured intracellular levels are lower than the calculated cell levels (T-EX) and the differences can be found back in the methanol solutions. In the samples that were processed at -20°C instead of -40°C, leakage was much more severe (A and B). Increasing the final concentration of methanol from 50% to around 90% was found to minimize leakage (D-G). At final methanol concentrations above 80% (F and G) most measured intracellular levels are within ±20% of the estimated cell levels (Figure 2.6). Further decreasing the temperature to -78°C did not result in any significant further improvement (H). In addition, leakage was most critical for smaller metabolites, which may permeate through the cell membrane more easily, while bulkier, more polar metabolites seemed to leak less. These results show that the standard protocol of quenching in 60% methanol at -40°C leads to considerable underestimation of the

intracellular levels of most metabolites. The results also highlight the importance of the temperature and the methanol concentration, two factors that were not fully explored before, in obtaining accurate data. Special care should be taken to ensure that the temperature is not significantly above -40°C throughout the sample treatment, since even at -20°C leakage may become much more severe. Furthermore, the results confirm that at least in the context of quenching, methanol seems to act as a cryopreservant rather than as an extractant and that increasing its final concentration to above 80% can effectively prevent losses of metabolites from the cells. These conclusions could not have been drawn and the quenching method could not have been reliably optimized without a comprehensive, fully quantitative approach.

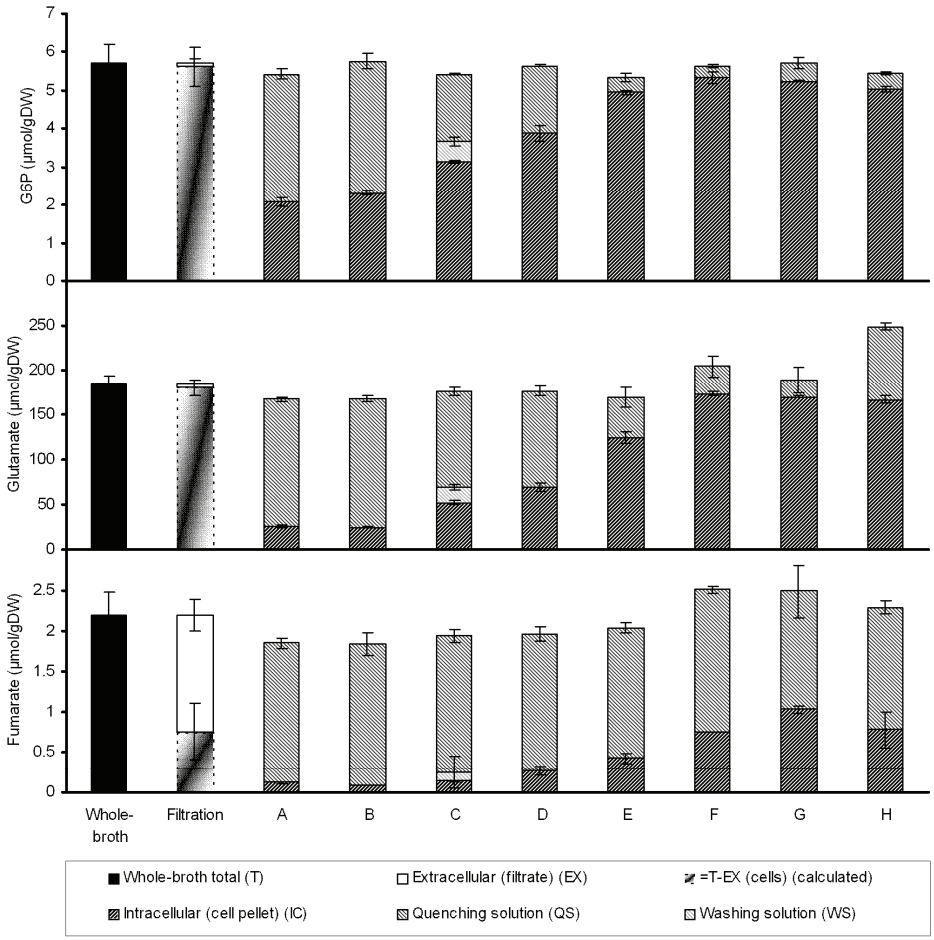


Figure 2.5: Full mass balance of the metabolite levels measured in the different sample fractions (see Figure 2.4) under 8 different variations in sample treatment protocol (see Table 2.1), exemplified for G6P, glutamate and fumarate. The metabolite levels inside the cells were estimated from the difference of the concentrations in whole-broth and extracellular medium (T-EX). Samples were from cultures at D=0.1 h-1 and 3.7 gDW/l. Data are averages ± standard deviation of at least 2 replicate samples, each analyzed at least in duplicate.

Table 2.2: Estimated cell levels (T-EX) and measured intracellular levels (IC) (see Figure 2.4), under 8 different protocol variations (see Table 2.1), for all metabolites analyzed. Samples were from cultures at D=0.1 h^{-1} and 3.7 g_{DW}/l. Values are averages ± standard error of at least 2 (IC) or 6 (T, EX) replicate samples, each analyzed at least in duplicate. Metabolites are in order of increasing molecular weight.

Metabolite	T-EX ($\mu mol/g_{DW}$)	IC ($\mu mol/g_{DW}$)							
		A	B	C	D	E	F	G	H
Glyoxylate	0.125 ± 0.056	0.032 ± 0.007	0.043 ± 0.006	0.041 ± 0.002	0.041 ± 0.011	0.043 ± 0.010	0.121 ± 0.018	0.094 ± 0.015	0.113 ± 0.013
Gly	3.18 ± 0.42	0.41 ± 0.01	0.43 ± 0.01	0.42 ± 0.01	0.63 ± 0.01	1.47 ± 0.05	2.49 ± 0.20	3.17 ± 0.12	3.03 ± 0.19
Pyruvate	1.35 ± 0.36	0.33 ± 0.01	0.42 ± 0.00	0.33 ± 0.01	0.70 ± 0.15	0.89 ± 0.16	1.19 ± 0.04	1.03 ± 0.07	1.04 ± 0.05
Ala	38.6 ± 0.6	3.1 ± 0.2	1.7 ± 0.0	3.6 ± 0.1	6.2 ± 0.3	16.1 ± 0.1	28.0 ± 0.4	34.3 ± 0.2	33.4 ± 0.3
Pro	4.60 ± 0.11	0.42 ± 0.02	0.27 ± 0.01	0.58 ± 0.03	0.95 ± 0.06	2.15 ± 0.04	3.42 ± 0.04	4.29 ± 0.03	4.02 ± 0.16
Fumarate	0.75 ± 0.09	0.12 ± 0.00	0.09 ± 0.00	0.14 ± 0.01	0.27 ± 0.03	0.42 ± 0.03	0.75 ± 0.00	1.03 ± 0.03	0.78 ± 0.16
Val	9.1 ± 0.3	1.2 ± 0.1	0.9 ± 0.0	2.0 ± 0.1	2.9 ± 0.2	6.4 ± 0.4	11.6 ± 0.9	10.3 ± 0.5	9.5 ± 0.4
Succinate	5.03 ± 2.16	0.32 ± 0.02	0.22 ± 0.00	0.21 ± 0.01	0.55 ± 0.08	0.94 ± 0.20	2.65 ± 0.29	4.42 ± 0.17	4.86 ± 0.20
Thr	4.42 ± 0.12	0.56 ± 0.02	0.37 ± 0.01	0.78 ± 0.04	1.20 ± 0.06	2.64 ± 0.03	3.92 ± 0.13	4.07 ± 0.04	3.92 ± 0.12
Leu	0.99 ± 0.09	0.22 ± 0.01	0.16 ± 0.00	0.28 ± 0.01	0.41 ± 0.03	0.73 ± 0.02	1.02 ± 0.05	1.01 ± 0.00	0.97 ± 0.03
Ile	1.71 ± 0.06	0.29 ± 0.01	0.20 ± 0.01	0.43 ± 0.02	0.65 ± 0.04	1.21 ± 0.03	1.68 ± 0.06	1.67 ± 0.00	1.60 ± 0.02
Asn	4.74 ± 0.15	0.56 ± 0.03	0.54 ± 0.02	1.37 ± 0.06	1.80 ± 0.08	3.75 ± 0.14	4.77 ± 0.09	4.75 ± 0.01	4.47 ± 0.02
Orn	4.18 ± 0.28	0.99 ± 0.05	1.08 ± 0.02	2.54 ± 0.03	2.86 ± 0.05	3.91 ± 0.09	4.31 ± 0.12	4.06 ± 0.07	3.97 ± 0.08
Asp	19.2 ± 0.5	2.0 ± 0.1	1.7 ± 0.0	3.9 ± 0.1	5.8 ± 0.5	13.2 ± 0.7	21.4 ± 0.0	20.8 ± 0.0	20.1 ± 0.2
Malate	6.27 ± 1.90	0.69 ± 0.04	0.56 ± 0.01	0.78 ± 0.00	1.52 ± 0.18	2.47 ± 0.20	6.09 ± 1.15	7.74 ± 0.52	7.94 ± 0.66
Oxoglutarate	2.30 ± 1.70	0.23 ± 0.01	0.18 ± 0.00	0.31 ± 0.01	0.76 ± 0.12	1.57 ± 0.02	1.61 ± 0.08	1.67 ± 0.01	1.99 ± 0.13
Gln	51.6 ± 2.1	5.9 ± 0.3	6.0 ± 0.2	15.3 ± 0.1	22.1 ± 3.2	45.2 ± 2.9	69.0 ± 1.8	65.9 ± 1.2	56.4 ± 0.2
Lys	3.84 ± 0.13	0.94 ± 0.04	0.85 ± 0.02	2.10 ± 0.04	2.65 ± 0.15	3.60 ± 0.22	4.04 ± 0.03	4.12 ± 0.07	4.03 ± 0.03
Glu	180 ± 2	26 ± 2	25 ± 0	52 ± 2	70 ± 2	124 ± 3	174 ± 2	170 ± 1	167 ± 3
Met	0.168 ± 0.037	0.035 ± 0.000	0.025 ± 0.001	0.033 ± 0.000	0.063 ± 0.010	0.105 ± 0.009	0.197 ± 0.004	0.208 ± 0.013	0.190 ± 0.011
His	5.23 ± 0.32	1.19 ± 0.04	1.39 ± 0.05	2.74 ± 0.12	3.87 ± 0.31	4.76 ± 0.49	6.29 ± 0.02	6.02 ± 0.07	5.84 ± 0.10
Phe	1.39 ± 0.11	0.22 ± 0.00	0.17 ± #DIV/0!	0.20 ± 0.00	0.50 ± 0.11	0.65 ± 0.28	1.55 ± 0.04	1.66 ± 0.07	1.64 ± 0.07
PEP	1.96 ± 0.14	0.05 ± 0.00	0.11 ± 0.00	1.19 ± 0.06	1.25 ± 0.07	1.93 ± 0.07	2.33 ± 0.02	2.31 ± 0.03	2.22 ± 0.01
G3P	0.176 ± 0.054	0.061 ± 0.003	0.059 ± 0.007	0.081 ± 0.012	0.088 ± 0.005	0.108 ± 0.006	0.113 ± 0.005	0.141 ± 0.005	0.139 ± 0.001
Tyr	1.84 ± 0.06	0.49 ± 0.02	0.48 ± 0.02	0.97 ± 0.04	1.18 ± 0.02	1.55 ± 0.05	1.53 ± 0.07	1.79 ± 0.07	1.58 ± 0.04
2PG+3PG	2.51 ± 0.09	0.17 ± 0.00	0.38 ± 0.00	1.42 ± 0.05	1.68 ± 0.06	2.35 ± 0.11	2.89 ± 0.00	2.77 ± 0.03	2.72 ± 0.06
Citrate	14.3 ± 3.3	4.3 ± 0.1	3.8 ± 0.2	4.9 ± 0.1	6.0 ± 0.2	7.1 ± 0.4	12.9 ± 0.2	16.8 ± 0.9	10.9 ± 0.1
Trp	0.577 ± 0.058	0.176 ± 0.005	0.151 ± 0.006	0.301 ± 0.012	0.406 ± 0.016	0.462 ± 0.022	0.526 ± 0.004	0.507 ± 0.021	0.502 ± 0.008
F6P	1.26 ± 0.05	0.30 ± 0.00	0.28 ± 0.00	0.57 ± 0.00	0.71 ± 0.02	1.29 ± 0.10	1.41 ± 0.02	1.46 ± 0.01	1.45 ± 0.00
G1P	1.46 ± 0.09	0.35 ± 0.02	0.35 ± 0.00	0.47 ± 0.02	0.67 ± 0.02	1.07 ± 0.02	1.34 ± 0.06	1.33 ± 0.03	1.24 ± 0.03
M6P	2.11 ± 0.13	0.67 ± 0.01	0.70 ± 0.01	1.37 ± 0.05	1.49 ± 0.08	1.89 ± 0.14	1.76 ± 0.07	1.78 ± 0.02	1.69 ± 0.01
G6P	5.62 ± 0.12	2.08 ± 0.07	2.33 ± 0.05	3.14 ± 0.02	3.89 ± 0.11	4.94 ± 0.02	5.33 ± 0.11	5.24 ± 0.00	5.03 ± 0.05
FBP	1.001 ± 0.231	0.584 ± 0.048	0.736 ± 0.021	0.562 ± 0.007	0.848 ± 0.074	0.731 ± 0.051	0.509 ± 0.022	0.755 ± 0.012	0.645 ± 0.003
T6P	0.379 ± 0.037	0.133 ± 0.002	0.218 ± 0.009	0.356 ± 0.007	0.380 ± 0.007	0.403 ± 0.005	0.458 ± 0.000	0.365 ± 0.021	0.344 ± 0.002

The occurrence of leakage in *S. cerevisiae* has also been reported by Villas-Boas *et al.* from measurements of metabolites in cell pellets, quenching solutions and filtrates [188]. In contrast, other studies have dismissed leakage mainly or exclusively because metabolite levels in methanol supernatants were below detection or considered negligible [35, 45, 72, 106]. Hans *et al.* actually found much higher levels of intracellular aminoacids after sampling into cold water (4°C) than after cold methanol quenching, which is a rather strong indication of leakage, but a mass balance was not performed [72]. To our knowledge, Gonzalez *et al.* were the only ones to report a mass balance (for 4 metabolites: G6P, ATP, NAD and NADH) including the measurement of metabolites in whole-broth to investigate leakage in *S. cerevisiae* [67]. They found that metabolite levels in methanol-quenched pellets were lower than in the whole-broth, particularly for G6P and NAD. However, the difference could be accounted for by the noticeably high amounts found extracellularly, which represented 45% of the total whole-broth G6P and 30% of NAD.

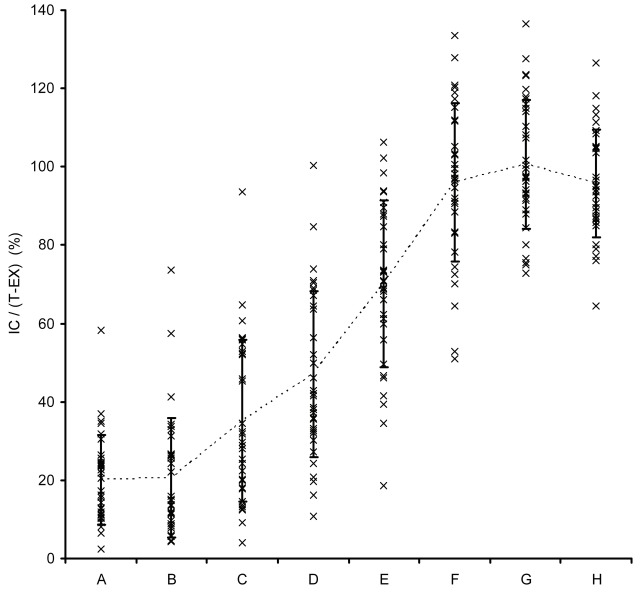


Figure 2.6: Ratios between measured intracellular level (IC) and estimated cell level (=T-EX) for each protocol variation (Table 2.1). Values are calculated from the data provided in Table 2.2. Each cross represents a metabolite. The dotted line joins the averages of all metabolites under each condition and the error bars represent the standards deviations

It should also be noted that all previous studies which in some way or another investigated the occurrence of leakage in *S. cerevisiae* were carried out with shake-flask cultures. Instead, the experiments described in this study were performed in aerobic glucose-limited chemostat cultures. The obvious advantage of chemostat cultures is that they are more reproducible [81]. In addition, while in batch cultures the time of sampling is critical because metabolite levels are highly dynamic [72, 195], in steady-state chemostat cultures metabolite levels change much more slowly or not at all [113, 208]. Furthermore, a larger fermenter volume and the use of rapid-sampling equipment mean that a large number of replicate samples can be obtained within a very short period of time and processed together, further improving data consistency and

reproducibility. The differences in growth conditions and sampling, as well as differences between *S. cerevisiae* strains and medium composition are likely to affect the biomass composition and sample properties, which in turn could affect the extent of metabolite leakage. This might explain, at least partly, some of the differences in results.

Our results show that the extent of metabolite leakage depends on the temperature during treatment and the properties of the quenching/washing solutions, namely the methanol concentration and the ionic strength. Throughout all experiments, including tryouts not shown here, the same trends were consistent and reproducible. To our knowledge, none of these factors had been quantitatively investigated before. One possible explanation for the effect of the methanol concentration might be that it decreases the freezing point (FP). The larger the difference between the processing temperature and the freezing point, the lower the chance that part of the sample freezes, causing cells lysis. However, we found that losses by quenching with 40% methanol (33% v/v final, FP≈-23°C [44]) and 60% methanol (50% v/v final, FP≈-42°C [44]) at -20°C were higher than with 60% methanol (50% v/v final, FP≈-42°C [44]) and 80% methanol (50% v/v final, FP≈-72°C [44]) at -40°C, respectively (Figure 2.5 A vs. D and B vs. E). Therefore, metabolite leakage depended both on temperature (Figure 2.5 B vs. D) and methanol concentration (Figure 2.5 D vs. E), but not on the difference between temperature and freezing point. In addition, if leakage occurred by cell lysis we would expect to find metabolites levels in the quenching/washing solutions in proportion to their levels inside the cells, which was not the case. What changes in methanol concentration, temperature and ionic strength have in common is that they can affect the solubility and diffusivity of metabolites. Methanol is a fairly polar solvent but it is by far not as good a solvent as water for hydrophilic compounds, so the higher its concentration the lower the solubility of most intracellular metabolites. On the other hand, increasing the ionic strength of a concentrated methanol solution may improve the solubility of polar compounds by allowing more ion-ion interactions. In addition, the solubility and diffusion coefficients usually increase with the temperature. In view of these considerations, we would expect higher methanol concentrations and lower temperatures and lower ionic strengths to reduce the solubility and diffusion rate of polar compounds, which is in agreement with the observed effects in terms of metabolite leakage from the cells. This is also in agreement with the fact that smaller metabolites were found to leak more than larger ones, as illustrated in Figure 2.7, since smaller metabolites are expected to have higher diffusivities.

Although increasing the methanol concentration could effectively prevent leakage of intracellular metabolites, this can also have one important disadvantage. We found that samples quenched and washed in solutions with higher methanol content contained much higher levels of phosphate and sulfate. Both are abundant in the medium used for the cultures (22mM phosphate and 40mM sulfate) and apparently were carried over in the samples, indicating that the washing efficiency was significantly decreased. While leakage can lead to underestimation of intracellular levels, insufficient washing can lead to overestimation. Extracellular components not adequately eliminated can also potentially interfere with certain analytical techniques (e.g. ion-suppression in MS-based analysis). Our results indicate that the intracellular levels were not overestimated even in samples quenched in pure methanol, since the measured IC levels were not consistently higher than the calculated cell levels (T-EX)

(Figure 2.6 F, G and H). However, in other culture conditions where extracellular concentrations are higher (e.g. batch, anaerobic conditions) this may need to be confirmed.

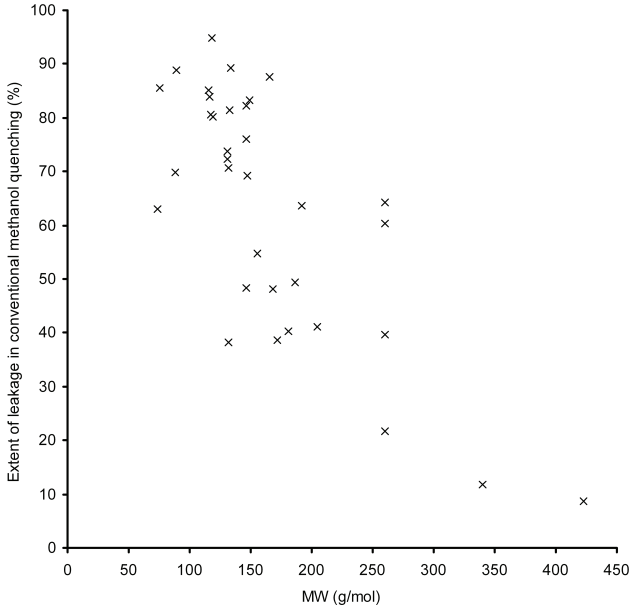


Figure 2.7: Relation between the molecular weight (proxy for molecular size) and the extent of leakage under the conventional cold methanol quenching (60% v/v, -40°C). The extent of leakage was calculated from the data in Table 2.3 (1 minus the inverse of the loss factor).

The importance of quenching metabolism, removing the extracellular medium and avoiding leakage

Although the need to stop metabolic activity upon sampling for intracellular metabolite analysis has been well documented for decades it is still possible to find studies being published which report intracellular metabolite measurements from samples obtained without any apparent concern for quenching. The turnover time, which is the ratio between the metabolite concentration and the flux through that pool, is a good criteria to evaluate the need for quenching. Since our results show that intracellular metabolite levels have been significantly underestimated due to leakage, we have re-calculated the turnover times based on our best estimate of the intracellular levels. These were obtained as the averages of the levels found with quenching in pure methanol (columns F, G and H of Table 2.2), where leakage was prevented (Figure 2.6). As shown in Table 2.3, most intermediates in central metabolism have short turnover times, particularly those that are highly connected (Pyruvate) or that have relatively low pools (FBP, fumarate). In this regard it is worth mentioning that cofactors such as ATP, ADP, NAD and NADH, which are involved in large numbers of reactions, are likely to have even shorter turnover times. On the other hand, most aminoacids have much longer turnover

times, owing to proportionally small fluxes and large pools. In view of these results we can conclude that sub-second quenching of metabolic activity is necessary for reliable analysis of intracellular levels of intermediates in central metabolism, while slower sampling methods such as filtration [16] may also acceptable for analysis of aminoacid levels.

Table 2.3: Intracellular levels, turnover times, extracellular fractions and loss factors due to leakage during conventional cold methanol quenching. Best estimates of the intracellular levels are the averages of levels obtained by quenching in pure methanol (average of columns F, G and H of Table 2.2). Net fluxes through metabolite pools were calculated from the estimated steady-state intracellular flux distribution under aerobic glucose-limited conditions according to [42]. The turnover time is the ratio between the intracellular level and the flux. The extracellular fraction is the ratio between the levels measured in the extracellular (EX) and total broth (T) samples. The loss factor due to leakage is the ratio between the best estimate of the intracellular level (in the second column) and the level found by conventional cold methanol quenching (column C of Table 2.3). Metabolites are in order of increasing molecular weight.

Metabolite	Intracellular level (μmol/g_{DW})	Estimated flux (μmol/g_{DW}/s)	Estimated turnover time (s)	Extracellular fraction	Loss factor due to leakage in conventional cold methanol quenching
Glyoxylate	0.11	*	*	70%	2.7
Gly	2.9	0.012	247	33%	6.9
Pyruvate	1.1	0.656	2	57%	3.3
Ala	32	0.010	3268	5%	8.9
Pro	3.9	0.004	925	8%	6.7
Fumarate	0.85	0.208	4	66%	6.2
Val	10	0.021	490	6%	5.1
Succinate	4.0	0.197	20	50%	19.2
Thr	4.0	0.018	220	9%	5.1
Leu	1.0	0.008	125	28%	3.6
Ile	1.6	0.012	140	15%	3.8
Asn	4.7	0.004	1142	9%	3.4
Orn	4.1	0.008	502	22%	1.6
Asp	21	0.036	577	4%	5.3
Malate	7.3	0.240	30	61%	9.3
Oxoglutarate	1.8	0.377	5	78%	5.6
Gln	64	0.027	2401	3%	4.2
Lys	4.1	0.007	619	8%	1.9
Glu	170	0.153	1112	2%	3.3
Met	0.20	0.003	66	42%	6.0
His	6.0	0.002	3141	10%	2.2
Phe	1.6	0.004	430	17%	8.1
PEP	2.3	0.404	6	8%	1.9
G3P	0.13	0.002	57	70%	1.6
Tyr	1.6	0.002	832	11%	1.7
2PG+3PG	2.8	0.422	7	8%	2.0
Citrate	14	0.228	59	50%	2.8
Trp	0.51	0.001	788	17%	1.7
F6P	1.4	0.197	7	2%	2.5
G1P	1.3	*	*	20%	2.8
M6P	1.7	*	*	1%	1.3
G6P	5.2	0.311	17	2%	1.7
FBP	0.64	0.197	3	6%	1.1
T6P	0.39	*	*	16%	1.1

*: Metabolites not present in the stoichiometric model [42] under aerobic glucose-limited conditions

As mentioned before, the main reason to use cold methanol quenching is that it allows removal of metabolites present in the extracellular medium. This is necessary if a significant fraction of the metabolites is present outside the cells. It may also be desirable in cases where medium components can interfere with analytical techniques (e.g. high salt concentrations can cause peak-shifting in ion-exchange LC and ion-supression in MS). If separation was not necessary, whole-broth extraction might be a simpler and faster alternative. Since we analyzed metabolite levels in all sample fractions, we can re-check the need to separate the extracellular medium by calculating the extracellular fraction for each metabolite (=EX/T). It should be kept in mind that a large extracellular fraction does not necessarily mean that the extracellular concentration is high. The reason for this is that the extracellular volume is typically 100 to 1000-fold higher than the intracellular volume so even very low extracellular concentrations can represent a large fraction of the total whole-broth amount (see Figure 2.5). As shown in Table 2.3, several metabolites have rather high extracellular fractions, particularly organic acids and aminoacids with non-polar side chains. Accurate analysis of the intracellular levels of these metabolites therefore requires adequate separation of the extracellular medium, while methods involving whole-broth extraction may be acceptable for analysis of compounds such as the phosphorylated intermediates, which are mostly present inside the cells. Note that these results were obtained in carbon-limited cultures and that extracellular fractions are most likely higher in carbon-excess conditions [16].

One possibility to differentiate between intracellular and extracellular metabolites is the "differential method", which we used above as benchmark for the improvement of the cold methanol quenching protocol. The major disadvantage of the differential method is the propagation of error associated with the subtraction, especially because metabolite levels in the filtrate are very dilute and difficult to assay reliably. This problem is further amplified if the extracellular fraction is large (two large numbers being subtracted to obtain a small number). The only way to try to minimize the uncertainty is through "brute force", by increasing the number of replicate samples (note that the uncertainties in Table 2.2 are standard errors, not standard deviations), but in practice this is sometimes not possible (e.g. investigation of short-term dynamics) and it is not guaranteed to succeed. The differential method is also not an alternative in the cases where medium components interfere with the analytical procedures.

It follows that the most desirable is to have a sampling method that ensures sub-second quenching of metabolism while allowing separation of the extracellular medium, which explains the popularity of the cold methanol method. Methanol is the most obvious choice of solvent because it is fully miscible with water (in contrast with more apolar solvents) but has a very low freezing-point (much lower than ethanol, propanol or glycerol, for example). In addition, methanol is thought to be less harmful to the cells than other organic compounds and methanol-water solutions are not very viscous, which allows easy centrifugation and cell separation. However, cold methanol quenching can only be used if leakage of intracellular metabolites does not occur. Much of the work reported here aimed at quantifying the extent of leakage and trying to prevent it. As shown in Table 2.3, the extent of leakage in the conventional cold methanol quenching method (60%, -40°C) is such that most metabolites investigated would be underestimated by more than 2-fold. With the exception of only the largest, most polar metabolites (FBP, T6P and probably others such as nucleotides), accurate

analysis of the intracellular metabolite levels in *S. cerevisiae* is only possible using the improved cold methanol quenching method described here.

These findings have obvious implications for metabolomics-based research in *S. cerevisiae*, such as non-stationary ^{13}C flux analysis, *in vivo* kinetic modeling and thermodynamic network analysis, which all rely on the accuracy of intracellular metabolite determination.

CONCLUDING REMARKS AND RECOMMENDATIONS

For accurate determination of intracellular metabolites in *S. cerevisiae* broth should be harvested into pure methanol at ≤-40°C with a sample/quenching solution ratio of 1:5 or lower (final methanol concentration ≥83%). Preferably, this should be done using rapid sampling equipment to avoid changes in metabolite concentrations during harvesting, especially from chemostat cultures and if short-turnover metabolites are to be measured. Centrifugation should be done quickly (≤5 min at ≥4000G) and using a pre-cooled rotor, to keep the temperature below or around -40°C. The overall time of exposure to methanol should be kept to a minimum. If a washing step is strictly necessary, to minimize the concentration of contaminants from the extracellular medium, this should be done with ≥83% methanol.

Sample treatment techniques should be validated quantitatively for each microorganism and growth condition. Not doing so compromises the conclusions drawn from subsequent research.

ACKNOWLEDGEMENTS

This work was funded by SenterNovem through the IOP Genomics initiative (project IGE3006A).

Chapter 3

Are we getting the right numbers?

A new, ^{13}C-based method for validation of sample treatment processes in microbial metabolomics

ABSTRACT*

Accurate determination of intracellular metabolite levels requires well-validated procedures for sampling and sample treatment. Several methods exist for metabolite extraction but the literature is contradictory regarding the adequacy and performance of each technique.

Using a strictly quantitative approach, we have re-evaluated 5 methods (Hot Water, HW, Boiling Ethanol, BE, Chloroform-Methanol, CM, Freezing-Thawing in Methanol, FTM and Acidic Acetonitrile-Methanol, AANM) for the extraction of 44 intracellular metabolites (phosphorylated intermediates, amino acids, organic acids, nucleotides) from *S. cerevisiae* cells. Two culture modes were investigated (batch and chemostat) to check for growth condition dependency and three targeted platforms were employed (2 LC-MS and 1 GC-MS) to exclude analytical bias. Additionally, for the determination of metabolite recoveries we applied a novel approach based on addition of ^{13}C-labeled internal standards at different stages of sample processing.

We found that the choice of extraction method can drastically affect measured metabolite levels, to an extent that for some metabolites even the direction of changes between growth conditions can be inverted. The best performances, in terms of efficacy and metabolite recoveries, were achieved with BE and CM, which yielded nearly identical levels for the metabolites analyzed. According to our results, AANM performs poorly in yeast and FTM cannot be considered adequate as an extraction method, as it does not ensure inactivation of enzymatic activity.

*Published as: Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics, in Analytical Chemistry (2009) 81, 17: 7379-7389

INTRODUCTION

The ability to accurately determine intracellular concentrations of metabolites is of key importance in studying signaling and metabolic reaction networks and their regulation *in vivo*. With the dawn of the metabolomics era, mainly fueled by the development of sensitive high-throughput MS-based analytics, metabolome-wide analysis is set to become reality. However, despite half a century of experience in the measurement of intracellular metabolites, the procedures used for preparation of the biological samples remain an issue [16, 17, 31, 166, 204]. It is increasingly recognized that, whether for semi-quantitative or quantitative purposes, the accuracy and reliability of the results are to a large extent determined by the first steps of sample treatment, namely rapid sampling, quenching of metabolic activity, separation of extracellular medium (when applicable) and metabolite extraction. We have recently addressed the problem of metabolite leakage from yeast cells during cold methanol quenching [31]. Here, we approach the main issues regarding the performance of extraction methodologies for microbial metabolomics, with a focus on *S. cerevisiae*.

There are three essential requirements that an ideal metabolite extraction method must fulfill [45, 67]: i) completeness of extraction, to ensure all intracellular pools are made entirely accessible for analysis, ii) prevention of metabolite conversion during the extraction or subsequent steps, mainly by ensuring effective inactivation of enzymes, and iii) absence of extensive degradation of metabolites by the procedure itself. With respect to this last requirement, partial losses may be acceptable if the results can be corrected by means of metabolite-specific recovery factors [45, 67] or through the use of adequate internal standards (e.g isotope-based) [13, 28, 209]. Theoretically, completeness of extraction cannot be directly verified, since one does not know *a priori* how much of each metabolite is initially present in the cells. Instead, completeness can be evaluated by comparing the ability of different methods to release metabolites from identical biological samples, which we shall refer to as efficacy. The absence of enzyme activity and the extent of metabolite degradation can both be tested by determining metabolite recoveries, traditionally by means of spiking or standard additions. Metabolite-specific recoveries much below 100% typically indicate degradation, while mixed recoveries above and below 100% for different metabolites would indicate inter-conversion.

Extraction can be achieved using high temperature, extreme pH, organic solvents, mechanical stress or combinations of these. Some of the most well-known extraction methods, such as perchloric acid (PCA) [71, 80], hot water (HW) [59, 207] and boiling ethanol (BE) [10, 58], have been employed at least since the 1950's. Eventually, the perceived need to adopt less aggressive procedures to prevent metabolite degradation seems to have steered researchers into the development of milder extraction methods. In the 1990's a neutral low-temperature technique based on a two-phase liquid extraction with Chloroform-methanol (CM) was introduced [45], presumably adapted from a method for lipid extraction from tissues [56]. More recently, two other methods have been proposed, one based on freezing-thawing cycles in methanol (FTM) [109], another using acidic acetonitrile-methanol (AANM) [141].

Table 3.1: Survey of comparisons of extraction procedures for intracellular metabolite analysis in microorganisms. * Legend: +, poor/bad; ++, fair; +++, good; underlined, best among the methods tested. Some variations in extraction time, buffers, solvent concentrations and temperatures are considered within the same method.

Source	Microrganism	Authors' evaluation, per extraction method*							Notes
		PCA	KOH	HW	BE	CM	FTM	AANM	
Bagnara and Finch, 1972 [7]	*E. coli*	+ + +		+					Quantitative, TLC, 6 metabolites (nucleotides), no recoveries. Tested 1 other method (TCA).
Lundin and Thore, 1975 [107]	5 bacteria species	+ +	+ + +	+ + +	+ + +				Quantitative, enzymatic analysis, 3 metabolites (ATP, ADP, AMP), recoveries by spiking. Tested 6 other methods.
Larsson and Olsson, 1979 [102]	4 algae species	+ + +		+ +	+ +				Quantitative, enzymatic analysis, 3 metabolites (ATP, ADP, AMP), no recoveries. Tested 3 other methods.
de Koning and van Dam, 1992 [45]	***S. cerevisiae***	+ + +				+ + +			Quantitative, enzymatic analysis, up to 13 metabolites (data not shown), recoveries by spiking (only CM)
Gonzalez et al, 1997 [67]	***S. cerevisiae***	+ +			+ + +				Quantitative, enzymatic analysis, 6 metabolites, recoveries by spiking
Hajjaj et al, 1998 [70]	*Monascus ruber*	+ +	+ +		+ + +				Quantitative, enzymatic analysis, 2-6 metabolites, recoveries by spiking
Hans et al, 2001 [72]	***S. cerevisiae***			+ + +	+ + +				Quantitative, HPLC, 17 metabolites (amino acids), recoveries by spiking (only BE)
Maharjan and Ferenci, 2003 [109]	*E. coli*	+	+		+ +	+	+ +		2D-TLC, semi-quantitative for efficacies (total extract intensity and relative intensities for 13 metabolites), no recoveries. Tested 1 other method (hot methanol)
Jernejc, 2004 [89]	*A. niger*	+ + +	+ + +		+ +				Quantitative, enzymatic analysis, 3 metabolites (organic acids), recoveries by stability in solution
Villas-Boas et al, 2005 [188]	***S. cerevisiae***	+	+		+	+ +	+ +		GC-MS, qualitative for efficacies (number of peaks detected), quantitative for recoveries, 27 metabolites, by spiking
Wang et al, 2006 [193]	*E. coli*		+ +		+		+ + +		Qualitative, ESI-MS (richness and reproducibility of mass spectra), no recoveries
Hiller et al, 2007 [77]	*E. coli*	+		+ + +	+ +				Quantitative, enzymatic analysis, 7 metabolites, recoveries by standard additions
Rabinowitz and Kimball, 2007 [141]	*E. coli*					+	+	+ + +	LC-MS, semi-quantitative for efficacies (peak heights for nearly 100 metabolites) and recoveries, by spiking (peak heights for 12 metabolites). Tested 2 other methods (and some variations).
Faijes et al, 2007 [52]	*Lactobacillus plantarum*	+ + +			+ + +	+ +	+ + +		Quantitative, enzymatic analysis, 3 metabolites, recoveries by spiking (only FTM). Tested 1 other method (chloroform-water)
Winder et al, 2008 [201]	*E. coli*	+	+		+ + +	+ + +	+ + +		GC-MS, qualitative (number of peaks detected) and semi-quantitative (relative abundance of 21 metabolites) for efficacies, no recoveries

Given the variety of extraction methods available, how does one decide which method to use? To demonstrate the complexity of this choice we have assembled a complete (to the best of our knowledge) survey of method comparisons in the literature, which we present in Table 3.1. As this overview shows, each of the methods is considered very good, or best, by some studies and poor, or inadequate, by others (with the exception of AANM which was evaluated only once). Since method comparisons have been performed in different species, when comparing them one should keep in mind that some aspects of extraction performance can be organism-dependent. In particular, completeness of extraction may be microbe-specific, since some types of cells are more susceptible to lysis than others (e.g. eukaryotes vs. prokaryotes, gram-positive vs. gram-negative). Enzyme resistance to inactivation may also be organism-dependent, although among mesophile species one would expect it to be comparable. As for metabolite stability, it should be nearly independent of the type of cell extracted. Although these factors must be considered, they seem unlikely to change the performance of a particular method from bad to best, nor can they explain the discrepancies found within the same, or closely related species. Another observation from the overview in Table 3.1 is the prominence of mild extractions in recent literature. Five out of the last 8 studies consider FTM the best extraction method. Although cold methanol is known to release intracellular metabolites [31, 105, 204], it is less clear whether enzyme activity can be effectively inactivated, especially if one considers that enzyme activities can be assayed at temperatures below 0°C in 70% methanol [19, 120]. Furthermore, of the 15 studies examined, 5 did not determine metabolite recoveries, without which it is not possible to check for metabolite conversion or degradation, and 8 investigated only 2-3 extraction methods, which limits the conclusions that can be drawn in terms of efficacy. In addition, all but 3 studies examined relatively small sets of compounds of very similar molecular properties. Finally, 5 of the last 8 studies based their evaluation on qualitative or semi-quantitative measures. Qualitative measures (e.g. number of GC peaks) bear little or no information on the fulfillment of the essential requirements, while the use of semi-quantitative measures (e.g. peak area or height, normalized intensities) generally implies the assumption of linearity of response and absence of matrix effects, which often is not valid for MS-based analysis of complex samples such as cell extracts [28].

The purpose of this study was to carry out an objective and quantitative re-evaluation of available extraction methods. The model organism *S. cerevisiae* was used, for which we have previously validated sampling and quenching procedures [31, 100]. Of the techniques in Table 3.1 only acid (PCA) and base (KOH) extractions were not included, because their salt-rich extracts present limitations for our analytical platform (e.g. ion suppression). These methods have also seen relatively little use in recent years, due to reported near-total losses of specific coenzymes [67, 70, 188]. We have based our evaluation only on strictly quantitative data of metabolite concentrations in cell extracts. Samples were obtained under well-defined, reproducible culture conditions using optimized sampling procedures. Two different growth states were investigated (glucose limitation and excess), to check for growth condition dependency. Three analytical methods (GC-MS and 2 LC-MS/MS) were employed, to exclude analytical bias, covering a total of 44 metabolites of different classes and molecular properties (phosphorylated intermediates, organic acids, amino acids, nucleotides). For the determination of metabolite recoveries we introduced and applied a novel approach based on differential addition of ^{13}C-labeled internal standards at the beginning and end of the sample treatment.

We are confident that the outcomes of this work will serve researchers in the microbial metabolomics community by providing an improved assessment of the extraction techniques at their disposal. The experimental strategy employed, including the new ^{13}C-based approach for determining recoveries, should also be useful as a framework for systematic validation of these, and newer methods, in other microorganisms.

EXPERIMENTAL PROCEDURES

Solvents and chemicals – HPLC-grade solvents were from J.T. Baker. Analytical grade standards were from Sigma-Aldrich.

Strain and medium – The strain used was the prototrophic haploid *Saccharomyces cerevisiae* CEN.PK 113-7D (MATa)[176]. Strain stocks were kept in 20% glycerol at -80°C. A new low-salt minimal medium was designed, for sustaining a biomass concentration of 1.5-4.5 g_{DW}/L. Ensuring low residual concentrations of salts minimizes interferences in the analysis (e.g. peak-shifting, ion suppression) by ions carried over in samples, in case of limited washing [31]. No change in morphology or physiology has been observed in the cells compared to growth on standard salt-rich medium [138, 185]. The composition of the optimized medium was, per L: 0.3 g $(NH_4)_2SO_4$, 3 g $NH_4H_2PO_4$, 0.3 g KH_2PO_4, 0.5 g $MgSO_4.7H_2O$, 1 g vitamin solution [24] and 1 g trace element solution [138]. As carbon source the medium contained 7.5 g/L glucose and as antifoam it contained 0.05 g/L silicone antifoaming agent (BDH, UK). The final pH was adjusted to 5 with KOH.

Cultivation conditions – Cultivations were performed aerobically in a 7 L fermentor (Applikon, The Netherlands) equipped with a DCU3 controller and continuous data acquisition via MFCS (B. Braun Biotech, Germany). A working volume of 4 L was maintained by weight control. The cultivations were carried out at 30°C, pH 5.0 (via addition of 2 M KOH), with 0.3 bar overpressure, an aeration rate of 0.5 vvm (120 L/h) and a stirrer speed of 500 rpm. Base utilization and effluent accumulation were monitored by weight. Oxygen and CO_2 fractions in dry exhaust gas and air were measured using a combined paramagnetic/infrared analyzer (NGA 2000, Rosemount, USA) via a MUX multiplexer unit (B. Braun Biotech, Germany). Pre-cultures were carried out in unbaffled 500 ml conical flasks on 100 ml minimal medium as described above, adjusted to pH 6, inoculated with 2 ml strain stocks. These were grown overnight in a rotary shaker, at 30°C, 200 rpm, and then used to inoculate the fermentor. Two growth conditions were investigated: chemostat and batch. The glucose-limited chemostat culture was carried out at a dilution rate of 0.1 h^{-1}. Sampling took place after 6 residence times (60 h) of steady-state growth with constant dissolved oxygen and off-gas readings, at a biomass concentration of 3.7 g_{DW}/L. For the batch culture, the medium was supplemented before inoculation with a sterile concentrated glucose solution, for an initial concentration of 30 g/L. Sampling took place in exponential glucose-consumption phase, 11 h after inoculation, at a biomass concentration of 3.2 g_{DW}/L. Biomass dry weight concentration was determined gravimetrically on membrane disk filters (PES, 0.45 µm, Pall, USA) dried for 48 h at 70°C. Pre-weighed filters were loaded with 5-10 g broth and washed with 3 volumes of demineralized water.

Samples for intracellular metabolites – For an overview of the experimental design, see Figure 3.1. Samples were taken using a custom-made low dead-volume rapid sampling setup [100] and leakage-free quenching was performed according to Canelas *et al* [31]. Briefly, 1.2 g (±0.06) of broth was withdrawn (<1 s) into 5 ml pure methanol at -40°C, the mixture was quickly vortexed (≈1 s) and placed back in the cryostat (-40°C) (Lauda, Germany). Extracellular medium was separated by centrifugation (5000G, 5min, -20°C, rotor precooled to -40°C). No additional washing step was performed. After decanting, 200 µl U-^{13}C-labeled cell extract was added to half of the samples as internal standard (IS) [117, 209]. Each of the 5 different extractions was performed in quadruplicate: two samples with ^{13}C-labeled IS (name AQ: after quenching) and two samples without (BA, see below) (Figure 3.1).

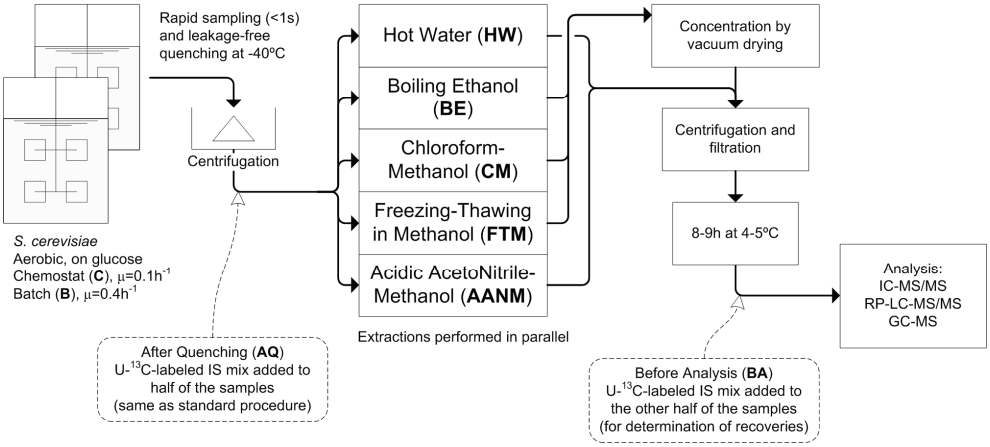


Figure 3.1: Schematic representation of the experimental design.

Metabolite extraction methods – The procedures for each extraction method were adapted, in some cases with minor simplifications, from the most relevant literature sources. For HW, BE, CM and FTM these were the protocols most commonly used for *S. cerevisiae*. In the case of AANM, for which there is yet no source in the literature referring its application to yeast cells, we employed the protocol that gave optimal results in *E. coli*. The details of each procedure are described individually below.

Hot water (HW) extraction – This procedure was adapted from [16, 72, 108, 111, 131]. Tubes containing 2 ml demineralized water were preheated in a water bath at 95-100°C for 6 min. Then, each sample was taken from the cryostat, the hot water was quickly poured over the cell pellet, the mixture was immediately vortexed and the sample was placed in the water bath. After 15 min each tube was placed on ice and the extracts were stored at -80°C until further use.

Boiling ethanol (BE) extraction – This procedure was adapted from [35, 67, 72, 100]. Tubes containing 5 ml 75% ethanol were preheated in a water bath at 95°C for 5 min. Then, each sample was taken from the cryostat, the boiling ethanol was quickly poured over the cell pellet,

the mixture was immediately vortexed and the sample was placed in the water bath. After 3 min each tube was transferred back to the cryostat (-40°C) and the extracts were stored at -80°C until further use.

Chloroform-methanol (CM) extraction – This procedure was adapted from [45, 162, 167, 188]. Throughout the procedure the temperature of samples and solutions was maintained as close as possible to -40°C. Each sample was resuspended in 2.5 ml precooled 50% (v/v) aqueous methanol, after which 2.5 ml precooled chloroform was added. All samples were then vigorously shaken for 45 min with an orbital shaker, using a custom-made tube adaptor, inside a -40°C freezer (temperature never above -32°C). The samples were then centrifuged (5000G, 5min, -20°C, rotor precooled to -40°C), the upper water/methanol phases were collected separately and the lower layers were reextracted with 2.5 ml precooled 50% (v/v) methanol by vortexing for 30 s. After centrifugation the upper phases were pooled with the first extracts and the combined extracts were stored at -80°C until further use.

Freezing-thawing in methanol (FTM) extraction – This procedure was adapted from [109, 187, 188]. Each sample was resuspended in 2.5 ml 50% (v/v) aqueous methanol precooled to -40°C, then frozen in liquid nitrogen for 3-5 min and thawed on ice for 3-5 min. After three freeze-thaw cycles the samples were centrifuged (5000G, 5min, -20°C), the supernatants were collected separately, and the pellets were reextracted with 2.5 ml precooled 50% (v/v) methanol by vortexing for 30 s. After centrifugation the supernatants were pooled with the first extracts and the combined extracts were stored at -80°C until further use.

Acidic acetonitrile-methanol (AANM) extraction – This procedure was adapted from [141]. Each sample was resuspended in 1 ml precooled (-20°C) acidic acetonitrile/methanol/water (40:40:20 v/v, containing 0.1 M formic acid) solution and placed in a cryostat at -20°C for 15min. The samples were then centrifuged (5000G, 5min, 4°C), the supernatants were collected separately and the pellets were reextracted twice with 200 µl ice-cold AANM solution, each time by vortexing for 20 s. Finally, the pooled extracts (1.4 ml) were neutralized with 100 µl 1.4 M NH_4OH and stored at -80°C until further use.

Sample concentration and conditioning – After thawing, BE, CM and FTM extracts were evaporated under vacuum (120 min, 30°C, <10 mbar) using a RapidVac (Labconco, USA). Dried residues of samples with ^{13}C-labeled IS (AQ) and without (BA) were resuspended in 600 µl or 400 µl milliQ water, respectively. All samples, including the HW and AANM extracts, were then centrifuged (15000G, 5 min, 4°C) to remove cell debris. The supernantants were transferred to filter caps (Ultrafree, PVDF, 0.22 µm, Millipore), centrifuged again and the filtrates were collected and stored at -80°C. Finally, to simulate the effect of typical waiting times during an analysis run, all samples were thawed and left to stand in the refrigerator (4-5°C) for 8-9 h. At the end of this period, 200 µl ^{13}C-labeled IS was added to the half of the samples which had not received it yet (named BA: before analysis) and all samples were placed back at -80°C until analysis.

Metabolite analysis – The concentrations of glucose-6-phosphate (G6P), fructose-6-phosphate (F6P), glucose-1-phosphate (G1P), mannose-6-phosphate (M6P), fructose-1,6-bisphosphate (FBP), trehalose-6-phosphate (T6P), 6-phospho-gluconate (6PG), 3-phosphoglycerate + 2-phosphoglycerate (3PG), phosphoenolpyruvate (PEP), glycerol-3-

phosphate (G3P), Citrate + Isocitrate (Citrate), Oxoglutarate, Succinate, Fumarate, Malate, sedoheptulose-7-phosphate (S7P) and UDP-Glucose were determined by anion-exchange LC-MS/MS [172]. The concentrations of AMP, ADP, ATP, UMP, UDP, UTP, GDP and CDP were determined by ion-pair reverse-phase LC-MS/MS [158].

The concentrations of 19 amino acids were determined by GC-MS using the EZ:Faast Free Amino Acid Kit (Phenomenex, USA). Extraction and derivatization of the amino acids was carried out according to the instructions of the manufacturer. Briefly, after a solid-phase extraction (SPE) step the amino acids are simultaneously derivatized in an aqueous medium on the amino- and carboxyl group with n-propyl chloroformate in the presence of n-propanol and 3-picoline, resulting in n-propoxycarbonyl n-propylesters. Owing to the SPE step and the high specificity of the derivatization reagent propyl chloroformate, sample matrix effects as high salt and buffer concentrations are limited and other components as sugars and organic acids do not interfere with the measurement. The derivatized extracts were analyzed with a Trace GC Ultra equipped with a programmed temperature vaporizer (PTV) injector and autosampler AI 3000 which was directly coupled to a Trace DSQ single quadrupole mass spectrometer with an electron ionisation source (Thermo Finnigan USA). One microliter of derivatized sample was injected on a ZBAAA column (medium polarity fused silica capillary column 10 m × 0.25 mm ID with a film thickness of 0.25 μm). The PTV injector, equipped with a glass liner with, carbofrit, was set in PTV-splitless mode with an initial temperature of 45 °C and a splitless time of 1 min. The temperature of the PTV was raised at 14.5 °C/s to 280 °C, held for 5 min, and subsequently raised at 14.5 °C/s to 300 °C. The GC temperature was initially set at 45 °C and then raised at 30 °C/min to 320 °C and held for 1 min. Helium was used as carrier gas with a gas flow of 1.5 mL/min. The transfer line to the MS was set to 250 °C and the ion source to 240 °C. The electron ionization was operated with 70 eV. For quantitative measurements, the MS was used in selected ion monitoring (SIM) mode.

All analyses were performed in duplicate. U-^{13}C-labeled cell extract was added as internal standard to both samples and calibration standards for IDMS-based quantification [117, 209].

RESULTS AND DISCUSSION

Whether for quantitative or qualitative purposes, the choice of extraction method can affect the results

An objective evaluation of intracellular metabolite extraction methods must be based primarily on the fulfillment of three essential criteria: completeness of extraction, prevention of inter-conversion (e.g. due to enzymatic activity) and absence of extensive degradation (e.g. chemical, thermal). Here we have evaluated these criteria using a strictly quantitative approach, that is, based on the absolute determination of metabolite levels in cell extracts obtained using different extraction techniques.

In these experiments we have employed a novel approach for determination of metabolite recoveries, based on the addition of ^{13}C-labeled IS at the beginning and end of the sample

treatment (see Figure 3.1). For samples named AQ, ^{13}C-labeled IS was added right after quenching (as is standard procedure), while for samples named BA, it was added just before analysis. Thus, in BA samples the ^{13}C-labeled IS will only correct for analytical artifacts (e.g. sample matrix effects), while in AQ samples it will also compensate for losses during sample treatment (e.g. volume losses, partial degradation). Levels in AQ samples provide a measure of the ability of each extraction method to release metabolites from the cells, independent of their stability during treatment, that is, its efficacy. And by comparing levels in BA and AQ samples we can determine the change in concentration of each metabolite over the entire sample treatment, that is, its recovery.

In this respect, note that we are determining an overall process recovery, including extraction, concentration and a conditioning period at 4-5°C which was meant to simulate the effect of the time samples spend waiting for analysis (see Figure 3.1). The reason for doing so is that the extraction method applied determines the extent of changes in metabolites not only during the extraction itself, but also potentially further down the sample treatment until the actual point when the sample is injected in the LC or GC (for example, if enzymatic activity is not fully inactivated). By determining overall process recoveries, all such changes are taken into account.

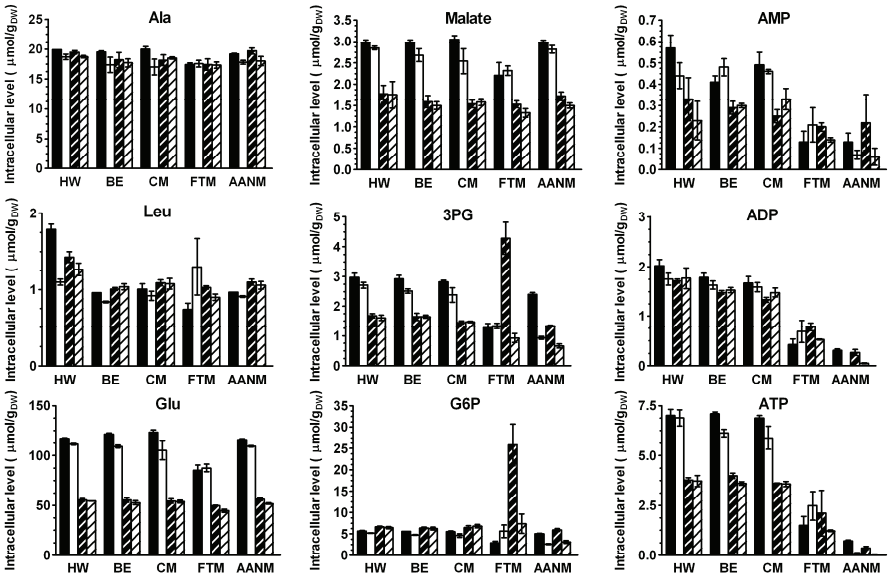


Figure 3.2: Intracellular levels of representative metabolites, obtained by extraction of cells from 2 growth conditions using 5 extraction methods (see Figure 3.1 for the experimental design). Data are averages and standard deviations of duplicate samples each analyzed twice. Legend: dark, AQ samples; light, BA samples; solid, chemostat; striped, batch. Comparison of dark and light bars indicates the effect of the sample treatment procedure, including the extraction, on the metabolite levels. Comparison of the solid and striped bars shows the differences in metabolite levels related to the growth condition.

In Figure 3.2 we present the measured metabolite levels in samples extracted with each of the 5 methods tested, under the two growth conditions investigated (chemostat and batch), for a set

of 9 representative metabolites (of the 44 analyzed) chosen to cover a wide range of molecular weight (from <100 to >500 Da), abundance (from <1 to >100 μmol/g_{DW}), functional class (amino acids, organic acids, phosphorylated intermediates, nucleotides) and analytical method used (3 each). From these data it is clear that different extraction techniques can indeed result in substantial differences in the measured metabolite levels. Overall, the CM, BE and HW methods seem to yield comparable metabolite levels, while AANM stands out with somewhat lower levels for the phosphorylated compounds and FTM can result in both higher and lower levels, depending on the metabolite and growth condition. Furthermore, even for experiments where the absolute metabolite levels are not required but only a qualitative impression of their changes, the choice of extraction method can still drastically influence the results. This can be seen for Leu, 3PG, G6P and the nucleotides in Figure 3.2, for which even the direction of change observed between growth conditions (glucose excess vs. glucose limitation) depends on the extraction technique used.

Comparison of extraction efficacy

As a measure of a method's ability to release metabolites from the cells we have defined extraction efficacy, versus the maximum among the five techniques, for metabolite x under extraction method i, as:

$$efficacy(vs.\,max)_{x,i} = \frac{[x]_i^{AQ}}{max[x]^{AQ}} \qquad \text{(equation 3.1)}$$

However, it became clear that in certain situations an extraction method might yield higher metabolite concentrations than the others if, for example, remaining substrate is converted to intermediates or macromolecules are broken down, in which case the efficacy of all other methods would appear, by comparison, misleadingly low. If so, an alternative measure would be to calculate an efficacy versus the median, rather than the maximum, which we defined as:

$$efficacy(vs.\,median)_{x,i} = \frac{[x]_i^{AQ}}{median[x]^{AQ}} \qquad \text{(equation 3.2)}$$

Note that these efficacy factors are comparative measures of effectiveness, so the results do depend on the set of techniques investigated. However, by comparing a sufficiently varied set of methods, based on different extraction principles and separately optimized for optimal release of metabolites, it seems safe to assume that at least one of them, if not more, will completely release any given metabolite. In this case, efficacies may be treated as near-absolute measures.

The extraction efficacies of all 44 metabolites analyzed, for both experiments, are plotted, for each extraction method, in Figure 3.3. The efficacy profiles for HW, BE and CM are all very similar, with efficacies near 1 for most of the metabolites. The profile for AANM is similar to the former for a wide range of metabolites but shows a clear discrimination of the larger, more polar metabolites (e.g. di- and tri-phosphate compounds), for which efficacies drop

significantly. Finally, the profile for FTM is the most scattered, with efficacies below 1 for a large number of metabolites, as well as exceptionally high efficacies for a small subset of compounds in the Batch experiment. A clear bias is also observed, as with AANM, towards lower extractability of the larger, more polar compounds.

The uniformity of the results with HW, BE and CM strongly indicates that these three methods achieve near-complete extraction of all metabolites analyzed. The only remarkable exceptions are six metabolites (PEP, 3PG, G6P, F6P, G1P and M6P) which were found at much higher levels in FTM extracts in the batch experiment. Strikingly, the very same metabolites are found at comparatively lower levels in the chemostat experiment. Furthermore, in both growth conditions the FTM method achieved overall poor efficacies, so it seems implausible that the difference lies in metabolite extractability. Rather, the fact that all six metabolites are fast-turnover glycolytic intermediates and their levels are comparatively higher in the condition of glucose-excess, but not glucose-limitation, points at the possibility of metabolite formation from the substrate (the ^{13}C-labeled IS would not correct for that). Furthermore, the metabolites involved imply that at least the whole glycolysis is active. This hypothesis is strengthened by the evidences of more widespread enzyme activity in FTM extracts provided by the recoveries, which we shall discuss below.

Among the similarity between the results of HW, BE and CM, a few differences stand out upon closer inspection. The profiles for BE and CM are particularly close, while HW yielded somewhat higher levels of certain metabolites, especially Leu, Met, GDP, Succinate, Phe and Ile (>2 standard deviations), followed by Trp, Fumarate, ADP, Ser and Tyr (>1 standard deviation). In principle, higher efficacies mean that HW has a better ability to extract these metabolites. However, it is unclear why a few amino acids and organic acids, but not others, should be better released by HW than, say, BE. The occurrence of several amino acids among the largest differences suggested an alternative possibility: that the higher levels could originate from decomposition of macromolecules such as proteins. If protein hydrolysis occurred, the largest relative contribution would be for those amino acids which are most abundant in whole-protein compared to their free pool. If we calculate, for each amino acid, the quotient between its percentage in protein [101] and its percentage in the free pool, the highest are for Leu (31), Phe (22), Met (16) and Ile (13), followed by Gly, Trp, Tyr and Pro (5-10). The striking correspondence with the amino acids which are higher in HW samples strongly indicates that partial protein hydrolysis could indeed be the source. Protein is a major component of biomass ($\approx$0.4 g/g_{DW}). We estimate that the hydrolysis of as little as 0.2% of the cell protein would be sufficient to explain the measured differences in free pools. This also complicates experimental verification of amino acid release from purified protein, as it is possible that only a small number of labile proteins are degraded. It is conceivable that the other metabolites elevated in HW extracts also originate from macromolecular degradation, although the exact biomass components that would generate succinate, fumarate, GDP and ADP upon thermal hydrolysis remain to be elucidated.

Comparison of metabolite recoveries

As a measure of the stability of each metabolite throughout sample treatment with each extraction technique we have defined the overall process recovery for metabolite x under extraction method i as:

$$recovery_{x,i} = \frac{[x]_i^{BA}}{[x]_i^{AQ}} \qquad \text{(equation 3.3)}$$

Since we are determining an overall process recovery, non-specific losses are to be expected simply by volume loss during centrifugation, filtration, etc. In our samples, estimated losses of 30-60 µl would represent 5-10% (recoveries 90-95%). Thus, even allowing for measurement error, losses of more than 15-20% indicate metabolite degradation, especially if they are metabolite-specific. Conversely, mixed recoveries above as well as below 100% would mean that metabolite conversion, either enzymatic or chemical, is taking place during the sample treatment. Metabolite conversion will distort the metabolic "snapshot" that each sample represents. An extraction technique which does not effectively prevent it (e.g. by inactivating enzymatic activity) cannot be considered suitable for metabolomics research.

The overall process recoveries of all analyzed metabolites, for both experiments, are plotted, per extraction method, in Figure 3.4. The recovery profiles for HW, BE and CM are quite comparable, with recoveries near 1 across all metabolites. The profile for BE is noticeably the "flattest", indicating the near-absence of metabolite-specific losses and excellent reproducibility. With CM and HW only a few metabolite-specific effects were observed: in the case of CM the recovery was low for Asn (0.38, batch) and Orn (0.71-0.74), while in the case of HW, it was low for Leu (0.61, batch) and Met (0.54, chemostat).

Extraction with AANM led to much more pronounced losses. While the recoveries for smaller, less polar compounds were very reproducible and near 1, a few mid-MW metabolites had low recoveries: Orn (0.48-0.52), Lys (0.63-0.68), His (0.54-0.59), PEP (0.35-0.38) and 3PG (0.39-0.49). For the larger phosphorylated intermediates the recoveries were systematically below 60% and some nucleotides (e.g. GDP, CDP, UTP, ADP, UDP) must have been so affected that they could not be detected in some of the samples. What these metabolites have in common is that they correspond exactly to the positively-charged side chain-containing amino acids (the remaining, Arg, was not analyzed) and the phosphorylated compounds. Why they are lost, however, is unclear. Exposure to low pH is not very long (45-60 min total) and at low temperatures. Besides, both compound classes are thought to be reasonably stable under acidic conditions. The AANM and ammonia solutions were titrated against each other before use (as a check), so prolonged exposure to low pH due to inadequate neutralization can be excluded. An alternative explanation could be that the most polar, charged metabolites are not entirely soluble in the cold solvent-rich mixture and, if partially precipitated (e.g. with cell debris, trace elements), would be lost during centrifugation and filtration.

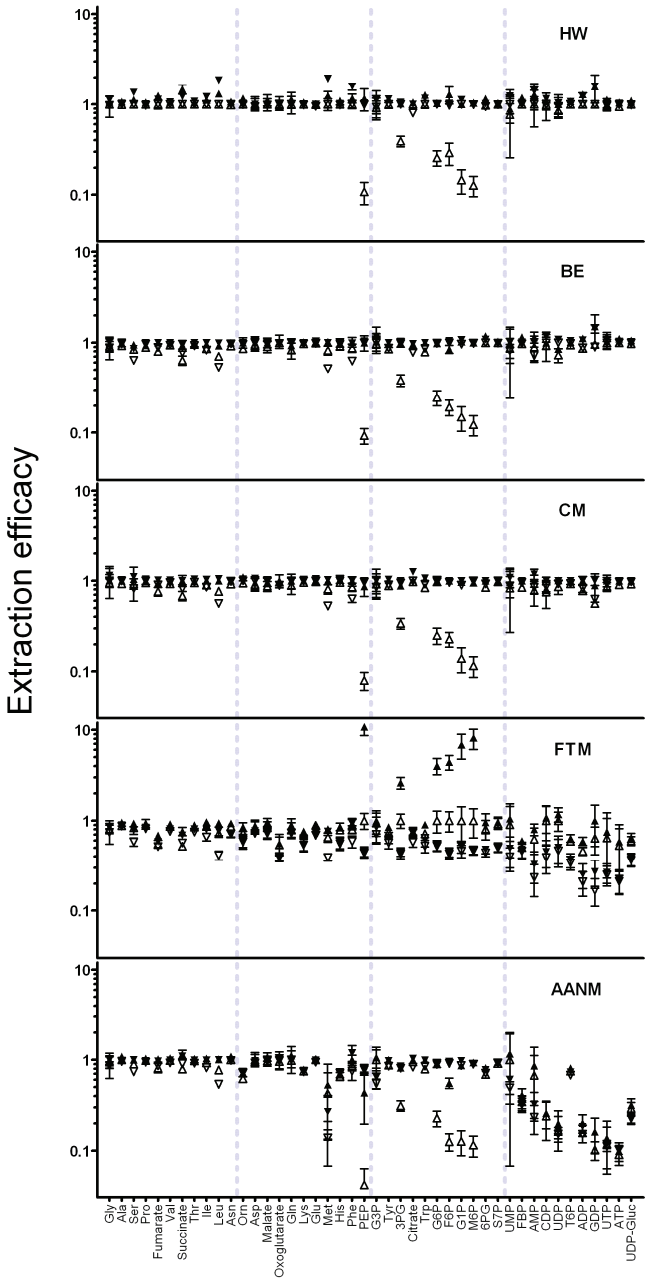


Figure 3.3: Extraction efficacies for all 44 metabolites analyzed, in order of increasing molecular weight, for each of the extraction methods, under both growth conditions (see Results for definitions of efficacy vs. max and efficacy vs. median). Data are averages and standard deviations of duplicate samples each analyzed twice. Legend: ∇ and Δ, efficacy vs. max in chemostat and batch, respectively; ▼ and ▲, efficacy vs. median in chemostat and batch, respectively. Gray dashed vertical lines are for guidance only.

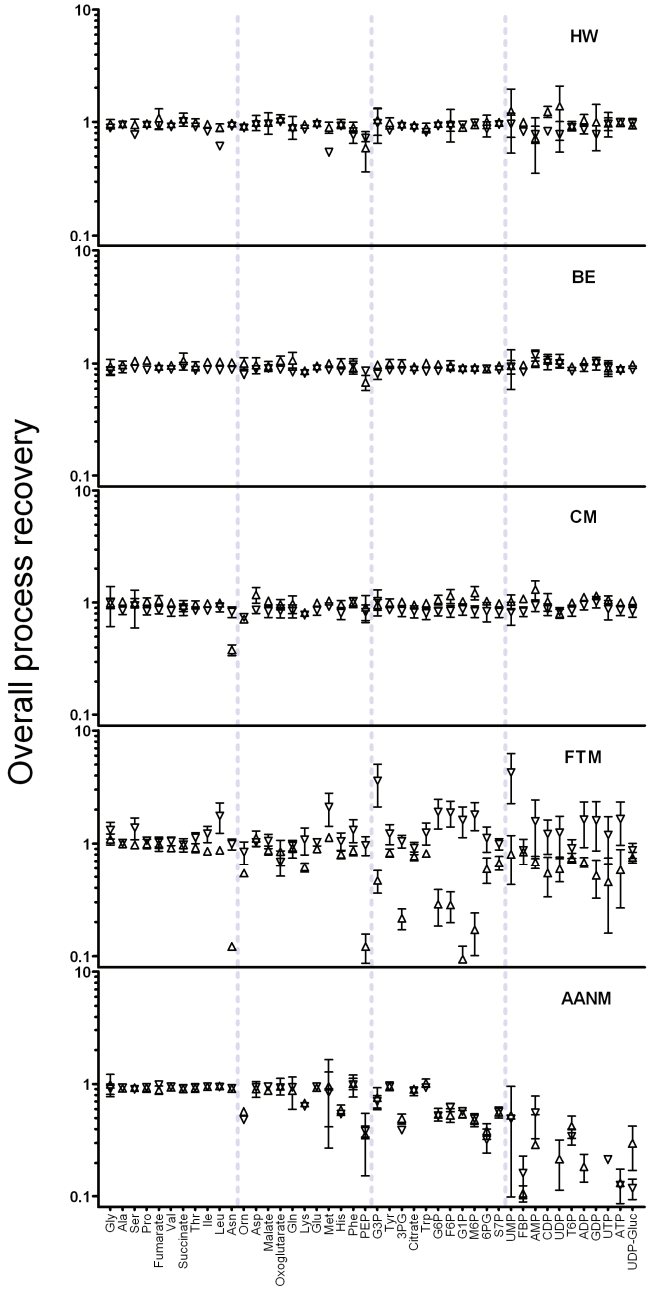


Figure 3.4: Overall process recoveries for all 44 metabolites analyzed, in order of increasing molecular weight, for each of the extraction methods, under both growth conditions (see Results for the definition of recovery). Data are averages and standard deviations of duplicate samples each analyzed twice. Legend: ∇, chemostat; Δ, batch. Gray dashed vertical lines are for guidance only.

The profile of recoveries for FTM extraction clearly shows that metabolite stability is most problematic with this method. Not only are there several observations of very low recoveries, there is also a large number of recoveries above 100%. This confirms the suspicions, discussed above, of remaining enzymatic activity during the procedure. In addition, rather than being reproducibly high or low in both experiments, for many metabolites the recovery depended on the origin of the samples. For example, Asn, Orn, Lys, PEP, 3PG, 6PG, S7P, T6P and UDP-glucose had low recoveries in samples from the Batch experiment, but not from the Chemostat, while Gly, Ser, Leu, Met and UMP had recoveries above 100% in samples from the Chemostat but not from the Batch. Furthermore, for a series of fast-turnover metabolites: G3P, G6P, F6P, G1P, M6P, AMP, CDP, UDP, ADP, GDP, UTP and ATP, recoveries were above 100% in Chemostat samples and well below 100% in Batch samples. Consider for instance the hexose-phosphates: in the Batch their levels in AQ samples were far higher than with the other extraction methods, which we have attributed to assimilation of glucose (see above). However, in BA samples their levels had decreased again, indicating conversion of the elevated pools into other intermediates later on in the sample treatment. This would have been partly compensated by the ^{13}C-labeled IS in the AQ samples, giving the very low recoveries, but not in BA samples. Conversely, in the Chemostat the hexose-phosphate levels in AQ samples were comparatively low, indicating poor extractability, but in BA samples the levels were higher, again indicating inter-conversion between metabolite pools after extraction. This could in principle occur at any stage after extraction, implying that enzymes were extracted from the cells but not actually inactivated. Additionally, it cannot be ruled out that conversion occurs also during extraction, since the samples spend several minutes thawing at 0°C and methanol and low temperatures do not prevent enzymatic activity but merely slow it down [19, 120]. While the latter could perhaps be prevented by thawing instead at -40°C, the former cannot. Note that the use of ^{13}C-labeled IS is not meant to deal quantitatively with metabolite conversion (that would require the substrate/product ratios in the IS to be exactly the same as in the sample), but rather with degradation and analytical artifacts (e.g. matrix effects). In conclusion, the changes observed in glycolytic intermediates are consistent with presence of enzymatic activity, and the extent of other metabolites exhibiting condition-dependency effects suggests this activity is widespread. Since inactivation of enzymatic activity is an essential requirement of an extraction method, these results mean that FTM cannot be considered a suitable technique for metabolomics research, at least in *S. cerevisiae*.

Finally, we would like to draw attention to the advantages of the approach used here for the determination of recoveries in comparison to the "traditional" approaches of spiking and standard additions. In principle, these involve additions (at one or multiple concentrations) of each metabolite to different samples. Furthermore, if the influence of sample treatment and analysis are to be distinguished, separate additions must be performed at different stages. Doing this separately for every metabolite requires so many samples that in practice mixes of metabolites are used instead. Not only does the addition of multiple metabolites make inter-conversion difficult to notice, the concentrations added are sometimes so high that they may no longer reflect what takes place in a real sample. This implies that the outcomes are dependent on the actual mix of metabolites chosen and concentrations applied. These limitation are circumvented by the ^{13}C-labeled IS approach designed here. Since the added (labeled) metabolites do not interfere with the (unlabeled) metabolites already present,

recoveries reflect changes as they actually occur in physiological samples. And since the IS mix is a cell extract itself, thus containing all metabolites, distinguishing metabolome-wide losses over the entire sample processing can be achieved with a modest experimental workload. Furthermore, if the recoveries at individual stages of sample treatment need to be distinguished, it is possible to extend the approach by applying the ^{13}C-labeled IS to additional replicate samples before and after each processing step. Although that will multiply the experimental effort several-fold, it should allow losses to be tracked down to specific operations.

Our verdict: best performance with BE and CM

From the quantitative data presented above, the 5 extraction methods investigated can be objectively evaluated in terms of completeness of extraction, prevention of metabolite conversion and metabolite stability. We conclude that the best extraction performance is obtained with the BE and CM methods. Both techniques achieved quantitative extraction of all metabolites analyzed with near-100% recoveries and good reproducibility. In the investigation of biological phenomena (RSD>15%) the metabolite profiles obtained with these two methods should be nearly indistinguishable, which is quite remarkable considering they are based on rather different extraction principles. The data from CM and BE samples have been combined and the resulting intracellular metabolite levels for the two growth conditions investigated are presented in Table 3.2.

Extraction with HW gave results very similar to BE and CM for most metabolites but raised some concerns regarding possible hydrolysis of macromolecules, which is undesirable. Although an incubation of 15 min is until now standard for extraction of yeast cells, it seems very likely that a shorter extraction time would lead to a performance comparable to that of BE. Extraction with AANM gave poor results, with limited efficacy and low recoveries for specific types of metabolites. Nevertheless, if the reason behind the problems encountered with the more polar, charged metabolites is found and solved, it could eventually provide a viable alternative to CM. Finally, according to our results FTM cannot effectively prevent metabolite conversion throughout the samples treatment, most likely by failing to inactivate enzymatic activity, and thus cannot be considered an adequate extraction method.

Considering the inconsistency in the literature concerning the evaluation of extraction methods illustrated in Table 3.1, it is not surprising that our conclusions are partly in agreement with some previous works and in disagreement with others. In the interest of clarifying possible causes for the differences in outcome, it is worth comparing our conclusions with those of previous studies. At first sight our most striking conclusion would appear to be that FTM is not adequate as an extraction method, since much of the recent literature, including some of the most influential papers on this topic, consider it the best technique [52, 109, 188, 193, 201]. However, three of these studies, including the original proponents of the method, did not determine metabolite recoveries [109, 193, 201], thereby leaving two of the essential requirements untested. In the study by Villas-Boas *et al*, metabolite recoveries were determined for 27 metabolites (14 of which also measured here) in extracts spiked with a mix of the same compounds [188]. Recoveries differed substantially between extraction methods and were

generally quite low, including for BE and CM extractions. Some metabolites, such as sugar phosphates, were not recovered acceptably (>30%) with any of the procedures, for which no explanation could be found. CM and FTM had, overall, the least unsatisfactory recoveries so the authors concluded FTM was the best choice, based on higher extraction efficacy (judged from the number of GC peaks detected) and lower solvent toxicity. A more recent study actually found evidence of conversion of ATP to ADP and AMP in ATP-spiked FTM extracts [52]. Despite this, the authors concluded that presence of enzyme activity was only a minor drawback of the method and, based on experimental simplicity and metabolite yields comparable to BE and PCA (in *L. plantarum*, for 3 metabolites), selected it as their preferred extraction method. From these considerations we conclude that, in the evaluation of the FTM method, the main sources of discrepancy between this work and the literature lie in the definition of evaluation criteria and the type of data used for their verification.

Our results also differ from the literature regarding the evaluation of the AANM extraction. The proponents of the method compared it to FTM and CM (and a few other variations) for the extraction of metabolites from *E. coli* [141]. Contrary to our results, they found that CM gave lower metabolite levels than AANM or FTM. However, the authors had adapted the procedure for CM extraction [45] by applying no mixing of the two-phase system. Interestingly, another study also in *E. coli* where only short periods of mixing (3 x 30 s) were applied for CM extraction also concluded it had limited efficacy [109]. Insufficient interfacial stress could well explain why CM failed to perform better than FTM and AANM, although it does not explain why it performed worse. This discrepancy may well be related to organism-specific differences. In the comparison between FTM and AANM, the authors preferred AANM because they found less metabolite conversion (attributed to chemical degradation) in nucleotide-spiked samples than with the FTM method. This is partially in agreement with our findings, since presence of enzymatic activity in FTM extracts (see above) would explain the conversions between metabolites which they observed. Finally, the authors also concluded that AANM gave the best performance for the extraction of nucleotides, which is opposite to our findings. If our hypothesis of metabolite precipitation were true, it is unclear why the *E. coli* samples would be less affected. On the other hand, it also cannot be excluded that losses were also taking place in their study since metabolite recoveries were not determined.

To our knowledge, BE, CM and HW extractions had not been compared together before, although several works compared at least two of them. Among those studies, half concluded they gave similar results [72, 102, 107, 201], while the other half found significant differences [52, 77, 109, 188]. As discussed above, in some instances the poor performance found with CM might be related to insufficient mixing of the two-phase system [52, 109]. Villas-Boas et al [188], also using *S. cerevisiae*, found particularly poor recoveries with BE extraction for several metabolites, among them hexose-phosphates (100% loss), Lys, Phe, Trp, Citrate and Oxoglutarate, which were completely stable in our experiments. The cause of this discrepancy is unknown. However, it should be noted that their finding of 0% recovery for hexose-phosphates is not in agreement with the fact that these metabolites can be routinely assayed in BE extracts at several labs [51, 106, 123, 197, 211]. In the study of Hiller et al [77], HW extraction was considered superior to BE in terms of reproducibility and linearity of standard additions (from which detection limits were determined). These aspects cannot be easily compared between studies since reproducibility

can be affected by so many different factors. In our study the reproducibility with HW and BE was comparable. In fact, the average RSD between metabolite levels (n=88) in duplicate biological samples was below 5% with either BE, HW or CM.

Table 3.2: Comparison of intracellular metabolite levels under the two growth conditions investigated.

Metabolite	Intracellular level (µmol/g_{DW})*	
	Chemostat (µ=0,1 h^{-1})	Batch (µ=0,4 h^{-1})
Glycolysis:		
G6P	5.53 ± 0.08	6.4 ± 0.4
F6P	1.38 ± 0.04	1.0 ± 0.1
FBP	0.47 ± 0.02	18 ± 1
3PG	2.9 ± 0.1	1.5 ± 0.1
PEP	2.1 ± 0.1	0.18 ± 0.03
Carbohydrate synthesis:		
G1P	1.07 ± 0.05	0.40 ± 0.02
UDP-Glucose	3.03 ± 0.09	0.60 ± 0.03
T6P	0.37 ± 0.02	0.080 ± 0.004
M6P	1.75 ± 0.03	0.72 ± 0.03
PPP and glycerol synthesis:		
6PG	0.69 ± 0.03	1.0 ± 0.1
S7P	2.97 ± 0.06	1.03 ± 0.03
G3P	0.09 ± 0.02	0.66 ± 0.05
TCA cycle:		
Citrate	7 ± 1	2.4 ± 0.1
Oxoglutarate	1.9 ± 0.1	2.8 ± 0.1
Succinate	1.09 ± 0.05	1.8 ± 0.2
Fumarate	0.655 ± 0.008	0.45 ± 0.03
Malate	3.01 ± 0.09	1.6 ± 0.1
Nucleotides:		
ATP	7.0 ± 0.2	3.8 ± 0.2
ADP	1.7 ± 0.1	1.40 ± 0.09
AMP	0.45 ± 0.06	0.27 ± 0.03
UTP	1.8 ± 0.2	1.2 ± 0.1
UDP	0.78 ± 0.05	0.38 ± 0.04
UMP	0.10 ± 0.02	0.06 ± 0.01
GDP	0.21 ± 0.04	0.13 ± 0.04
CDP	0.17 ± 0.02	0.10 ± 0.01
Amino acids - aromatic and His:		
Phe	0.47 ± 0.02	1.5 ± 0.1
Tyr	1.41 ± 0.05	1.01 ± 0.06
Trp	0.40 ± 0.02	0.190 ± 0.008
His	5.3 ± 0.2	7.7 ± 0.5
Amino acids - from 3PG:		
Ser	4.0 ± 0.9	12.4 ± 0.8
Gly	2.3 ± 0.4	12 ± 1
Amino acids - from pyruvate:		
Ala	19.8 ± 0.4	18 ± 1
Leu	0.98 ± 0.05	1.05 ± 0.05
Ile	1.72 ± 0.04	1.99 ± 0.09
Val	8.1 ± 0.1	8.9 ± 0.3
Amino acids - from oxoglutarate:		
Glu	122 ± 2	55 ± 2
Gln	45.1 ± 0.6	43 ± 5
Pro	3.9 ± 0.1	2.7 ± 0.2
Orn	3.9 ± 0.1	15 ± 1
Amino acids - from oxaloacetate:		
Asp	15.8 ± 0.3	12 ± 2
Asn	4.52 ± 0.04	2.7 ± 0.1
Lys	4.4 ± 0.1	20.2 ± 0.8
Met	0.218 ± 0.007	0.42 ± 0.02
Thr	4.1 ± 0.2	12.3 ± 0.7

* Data are averages and standard deviations of four samples, two extracted with BE and two with CM, each analyzed twice.

Finally, let us consider how our findings relate with the idea that comprehensive coverage of the metabolome can only be achieved using a combination of parallel extraction methods. This notion can be traced back to the works of Maharjan and Ferenci and Villas-Boas et al [109, 188]. Having observed large disparity in results with different extraction methods and in the absence of one method that was better for all metabolites, Maharjan and Ferenci concluded it was unlikely that a single extraction method could cover all metabolites, given their variety in chemical and physical properties. Their conclusion was based on relative measurements of 13 metabolites, most with poor reproducibility, and no data on metabolite recoveries. This frame of mind may have led Villas-Boas et al to express their results in terms of metabolite classes rather than individual compounds. Having observed large disparity in the recoveries of 27 metabolites, they concluded that every method discriminated particular metabolite classes and concluded that developing a method that could extract the large variety of compounds in the cell was an impossible task. Since then, these notions have been supported by at least two reviews on the topic (one of which by our group) [114, 128] and by some of the most recent literature [201]. An additional contributing factor may have been the inconsistency between published method comparisons, illustrated in Table 3.1. However, the results we have presented here are not consistent with that view. Admittedly, we have analyzed only a subset of the metabolome, from which is it dangerous to extrapolate. It cannot be excluded that extraction performances differ significantly for functional classes not analyzed here, such as lipids, sugars, CoA derivatives, etc. Nevertheless, the fact that all metabolites analyzed could be extracted with high efficacies and near-100% recoveries by at least two very different extraction techniques deserves being considered before assuming that metabolome-wide extraction is an impossible feat, in yeast as well as other microorganisms.

CONCLUDING REMARKS AND RECOMMENDATIONS

For accurate determination of intracellular metabolite levels in *S. cerevisiae*, samples should be extracted using Boiling Ethanol or Chloroform-Methanol. Both methods gave high efficacies and excellent recoveries for the whole range of metabolites analyzed here. Hot Water also performs generally well but raises the risk of over-estimating some metabolite pools, presumably due to macromolecule hydrolysis, although this can probably be minimized with a shorter incubation time. The performance of Acidic Acetonitrile-Methanol is inferior, for reasons not yet elucidated. Freezing-Thawing in Methanol is inadequate as an extraction method, as it does not ensure inactivation of enzymatic activity.

Some aspects of metabolite extraction are species-dependent so the performance of the different techniques available should be adequately validated with each microorganism. Furthermore, extraction performance may in some cases depend on the history of the sample, so method comparisons can benefit if investigations are performed under more than one growth condition (e.g. chemostat and batch).

The approach outlined here, of applying ^{13}C-labeled internal standards at different stages of the sample treatment process, is particularly useful for the determination of metabolite recoveries.

The choice of methods for the earliest steps in sample treatment can affect the outcome of a metabolomics-based experiment to the extent that even the direction of changes in metabolite levels may be distorted. Insufficient validation of the experimental procedures used thus compromises the results of subsequent research and conclusions derived thereof.

ACKNOWLEDGEMENTS

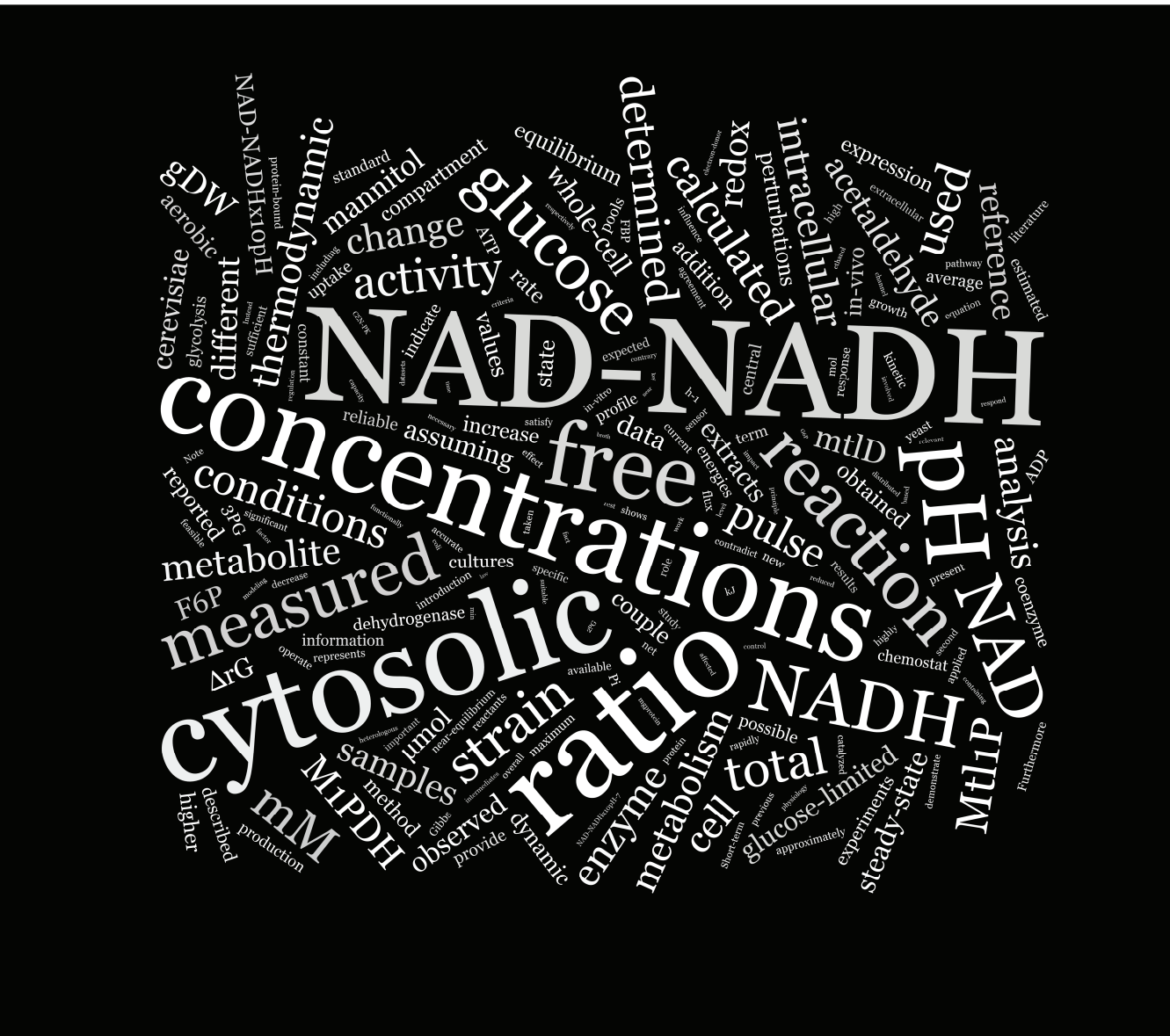
The authors are grateful to Sergio Rossel, Ana Luísa Cruz and Nathalie Bleijie for precious assistance in carrying out the extractions in parallel, Wouter van Winden and Aljoscha Wahl for insightful suggestions on the experimental design, Zhen Zeng for expert analytical work and Rob Kerste and Dirk Geerts for knowledgeable support in the fermentation lab.

This work was funded by SenterNovem through the IOP Genomics initiative (project IGE3006A) and carried out within the research program of Kluyver Centre for Genomics of Industrial Fermentation, which is part of NGI/NWO.

Chapter 4

Bypassing compartmentation

A new reporter-based method for determining the cytosolic free NAD/NADH ratio in *S. cerevisiae*



ABSTRACT*

The coenzyme NAD plays a major role in metabolism as a key redox carrier and signaling molecule but current measurement techniques cannot distinguish between different compartment pools, between free and protein-bound forms and/or between NAD(H) and NADP(H). Local free NAD/NADH ratios can be determined from product/substrate ratios of suitable near-equilibrium redox reactions but the application of this principle is often precluded by uncertainties regarding enzyme activity, localization and coenzyme specificity of dehydrogenases.

In *S. cerevisiae*, we circumvented these issues by expressing a bacterial mannitol-1-phosphate 5-dehydrogenase and determining the cytosolic free NAD/NADH ratio from the measured [fructose-6-phosphate]/[mannitol-1-phosphate] ratio. Under aerobic glucose-limited conditions we estimated a cytosolic free NAD/NADH ratio between 101(±14) and 320(±45), assuming the cytosolic pH is between 7.0 and 6.5, respectively. These values are more than 10-fold higher than the measured whole-cell total NAD/NADH ratio of 7.5(±2.5). Using a thermodynamic analysis of central glycolysis we demonstrate that the former are thermodynamically feasible, while the latter is not. Furthermore, we applied this novel system to study the short-term metabolic responses to perturbations. We found that the cytosolic free NAD-NADH couple became more reduced rapidly (timescale of seconds) upon a pulse of glucose (electron-donor) and that this could be reversed by the addition of acetaldehyde (electron-acceptor). In addition, these dynamics occurred without significant changes in whole-cell total NAD and NADH.

This approach provides a new experimental tool for quantitative physiology and opens new possibilities in the study of energy and redox metabolism in *S. cerevisiae*. The same strategy should also be applicable to other microorganisms.

* Published as: Determination of the cytosolic free NAD/NADH ratio in *S. cerevisiae* under steady-state and highly dynamic conditions, in Biotechnology and Bioengineering (2008) 100, 4: 734-743

INTRODUCTION

The compartmentation of metabolism in eukaryotes poses a problem for the accurate determination of relevant concentrations of intracellular metabolites. This is not sufficiently addressed by current "metabolomics" approaches, which provide total concentration averages over the entire cell. That is a major constraint to current efforts in quantitative physiology and kinetic modeling of metabolic reaction networks, which rely on the accuracy of the information regarding the *in vivo* environment surrounding the enzymes.

Thermodynamic analysis of glycolysis with published metabolite concentrations for *S. cerevisiae* reveals positive Δ_rG's, in disagreement with the 2nd law of thermodynamics (Table 4.1). The same contradiction can be found by thermodynamic analysis of datasets from other organisms such as *P. chrysogenum* and *E. coli* [99, 191]. At first sight, the data generated so far seems insufficient to understand even well-known pathways. As we shall demonstrate, this can mostly be accounted for by the lack of accurate data on the cytosolic free NAD/NADH ratio.

Table 4.1: Thermodynamic analysis of central glycolysis in yeast. Intracellular concentrations of glycolytic intermediates obtained from literature [82, 169, 170, 189]. All datasets refer to *S. cerevisiae* using glucose as main carbon source. Concentrations reported in $\mu mol/g_{DW}$ were converted using the factor 2,38 ml_{cell}/g_{DW} [170]. Gibbs energies of formation of the metabolites at reference pH'=7.0, T=25°C and I=0.25 M were obtained from [3]. The Gibbs energy of reaction was calculated for the overall reaction: FBP + 2 NAD + 2 ADP + 2 Pi → 2 3PG + 2 ATP + 2 NADH. A reaction is feasible if $\Delta_rG' \leq 0$.

Reference	Theobald et al, 1997[170]	Visser et al, 2004[189]	Teusink et al, 2000[169]	Hynne et al, 2001[82]
Strain and conditions	**CBS 7336, aerobic, glucose-limited, D=0.1 h^{-1}**	**CEN.PK 113-7D, aerobic, glucose-limited, D=0.05 h^{-1}**	**Compressed yeast (Koningsgist), anaerobic, glucose excess**	**X2180, glucose excess, oscillating**
FBP (mM)	0.11	0.17	5.51	5.1
3PG (mM)	1.33	0.67 [c]	0.9	0.3 [b]
ATP (mM)	3.36 [a]	2.65	2.52	2.1
ADP (mM)	0.47 [a]	0.72 [d]	1.32	1.5
Pi (mM)	9	43 [d]	10	11
NAD (mM)	1.34 [a]	1.81	1.2	0.65
NADH (mM)	1.09 [a]	0.12	0.39	0.33
NAD/NADH	1	15	3	2
Δ_rG' (kJ/mol)	+38	+10	+15	+10

a: estimated cytoplasmic concentrations for ATP, ADP, NAD and NADH were also reported. If those concentrations are used instead, the Δ_rG' is +21kJ/mol. b: not reported, taken from [147]. c: calculated from the concentration of 2PG+3PG assuming equilibrium of phosphoglycerate mutase, with $\Delta_rG'^{\circ} = 5.9$ kJ/mol [3]. d: not reported, taken from [210].

The coenzyme NAD is a key electron-carrier which mediates hundreds of reactions. The redox state of the NAD-NADH couple plays a central role in energy metabolism [8, 183]. It has also been shown to be involved in metabolic disorders [200], oxidative stress [48], cell ageing [104], signal transduction [11, 213] and transcriptional regulation [212]. NAD and NADH are present in the different cell compartments and their distribution is uneven. Furthermore, in addition to the

free species (kinetically and thermodynamically relevant), NAD and especially NADH are partly present in protein-bound form [26, 181]. Measurements of whole-cell, total concentrations of NAD and NADH do not distinguish between different compartment pools or between free and protein-bound forms, so they provide no information on the *in vivo* redox state of the NAD-NADH couple inside the cell's compartments. By fluorescence lifetime imaging it is possible to distinguish between free and protein-bound coenzymes, but not between NADH and NADPH [212], which have rather different metabolic roles.

These limitations can be overcome by calculating the NAD/NADH ratio from the concentrations of oxidized and reduced reactants of suitable near-equilibrium reactions:

$$\frac{[NAD]}{[NADH]} \approx \frac{[oxidized\ product]}{[reduced\ substrate].K'_{eq}} \qquad \text{(equation 4.1)}$$

In this way, one obtains the free NAD/NADH ratio in the compartment where the reaction takes place. However, this principle can only be applied if the following conditions are fulfilled:

- the enzyme catalyzing the reaction is strictly specific for its coenzyme
- the reaction takes place only in the compartment of interest (due to the localization of the enzyme, or the reactants, or both)
- the activity of the enzyme is high enough to establish near-equilibrium between the metabolite pools
- the equilibrium constant of the reaction is known
- the intracellular concentrations of all reactants (besides NAD and NADH) can be accurately measured
- those measured reactants are identically distributed in the cell (e.g. all cytosolic or all uniformly distributed)

In mammalian cells, the free NAD/NADH ratios in the cytosol and mitochondria have been determined from measured lactate/pyruvate and acetoacetate/β-hydroxybutyrate ratios, respectively [199]. This is possible because lactate dehydrogenase and β-hydroxybutyrate dehydrogenase are highly active in certain tissues such as liver and the conditions above are satisfied. In *S. cerevisiae*, as in many other organisms, the application of this principle is not straightforward. Lactate dehydrogenase and β-hydroxybutyrate dehydrogenase are not present. Alternatively, NAD/NADH ratios have been reported based on the metabolite ratios of alcohol dehydrogenase [79] or glyceraldehyde-3-phosphate dehydrogenase [170]. However, these reactions do not fully satisfy the criteria mentioned above. To our knowledge, there is no reaction in *S. cerevisiae* that does. Therefore, we proposed to introduce a heterologous enzyme especially for this purpose. We chose to use mannitol-1-phosphate 5-dehydrogenase (M1PDH), which catalyzes the following reversible reaction (EC 1.1.1.17):

$$Mtl1P + NAD \rightleftharpoons F6P + NADH + H^+ \qquad \text{(equation 4.2)}$$

Provided M1PDH can be functionally expressed in the cytosol of yeast at sufficiently high activity, it can satisfy all necessary criteria described above.

The standard equilibrium constant (K_{eq}) for the reaction catalyzed by M1PDH is defined as:

$$K_{eq} = \frac{[F6P].[NADH].[H^+]}{[Mtl1P].[NAD]} \qquad \text{(equation 4.3)}$$

This equation can be re-arranged to:

$$\frac{[NAD]}{[NADH]} \frac{1}{10^{-pH}} = \frac{[F6P]}{[Mtl1P].K_{eq}} \qquad \text{(equation 4.4)}$$

The right side of this equation contains only metabolite concentrations that can be measured and a standard equilibrium constant which is available from the literature: $7.9 \pm 0.6 \times 10^{-10}$ M [92]. The term on the left side of the equation gives us the redox state of the NAD-NADH couple, including the pH. The fact that the pH is included is a remarkable advantage, because the thermodynamic analysis of any reaction involving NAD(H) always requires information on the pH. By determining the ($NAD/NADH \times 10^{pH}$) ratio one circumvents the need to additionally determine the pH, which makes this principle ideal for thermodynamic reaction analysis.

Furthermore, if the cytosolic pH can be assumed or experimentally determined the NAD/NADH ratio can be calculated separately. Unfortunately, currently this may not always be possible because the techniques available to measure intracellular pH [1, 65, 94, 95] are not compartment-specific, thus providing an average cellular pH, and often require extensive handling of the cells, which is itself prone to affecting the cytosolic pH.

For practical reasons, it is common to express the equations above in terms of the apparent equilibrium constant (K'_{eq}) at specified reference conditions of pH (e.g. pH'=7.0), ionic strength (e.g. 0,25 M) and concentrations of ions such as magnesium, iron, phosphate, etc. In the case of the reaction catalyzed by M1PDH the pH is the most important factor and the previous equation can be written as:

$$\frac{[NAD]}{[NADH]} 10^{pH - pH'} = \frac{[F6P]}{[Mtl1P].K'_{eq}} \qquad \text{(equation 4.5)}$$

Note that for convenience each term has been multiplied by $10^{-pH'}$, but the term ($NAD/NADH \times 10^{pH-pH'}$) remains independent of the pH. At reference pH'=7.0, the K'_{eq} is $7.9 \pm 0.6 \times 10^{-3}$.

MATERIALS AND METHODS

Strains – The strain referred to as mtlD strain was obtained by transforming *Saccharomyces cerevisiae* CEN.PK 113-7A (MATa HIS3-Δ1) [49] with the plasmid pYX222:mtlD using the LiAc/SS carrier DNA/PEG method [64], at BIRD Engineering (Schiedam, The Netherlands). The plasmid pYXX222:mtlD is based on the yeast multicopy expression vector pYX222 (R&D

Systems Europe, UK) and carries *HIS3* for selection purposes and the *E. coli mtlD* gene under the control of a *TPI* promoter [41]. The strain referred to as reference strain is *Saccharomyces cerevisiae* CEN.PK 113-7D (MATa) [176].

Chemostat cultivations - The cells were grown in aerobic carbon-limited chemostat cultures in a 7 l fermentor (Applikon, Schiedam, The Netherlands) with a working volume of 4 l, on defined mineral medium [185] with 7.5 $g.l^{-1}$ glucose. The dilution rate was 0.1 h^{-1} and the aeration rate was 0.5 vvm (120 $l.hr^{-1}$). Dissolved oxygen tension (DOT) was measured in-situ with an oxygen probe (Mettler-Toledo) and pO_2 and pCO_2 in the off-gas were measured at-line using a combined paramagnetic/infrared analyzer (NGA 2000, Fisher-Rosemount). The pH was controlled at 5.0 with 4M KOH and the temperature was set at 30°C. The overpressure in the vessel was kept at 0.3 bar and the stirrer speed was 600 rpm, ensuring that the DO was always above 80%. The cultures were sampled after at least 3 volume changes of steady-state glucose-limited growth with stable DOT and off-gas measurements.

Short-term perturbation experiments - Short-term perturbation experiments were carried out using a miniaturized plug-flow reactor, the BioScope [116]. In brief, broth was withdrawn from the fermentor (at 1.34 ml/min), mixed with a concentrated pulse solution (0.16 ml/min) and flown through a serpentine-shaped channel. The pulse solution was either 21 mM glucose or 21 mM glucose + 21 mM acetaldehyde. This resulted in a 2.3 mM glucose pulse or a 2.3 mM glucose + 2.3 mM acetaldehyde pulse, respectively. Aeration was maintained through a gas channel that contacts the liquid channel. An automated sampling system controlled a series of valves to withdraw samples at different channel lengths, representing periods of exposure from 4 to 120 s.

Samples for intracellular metabolite analysis - For steady-state concentrations, samples were taken from the fermentor using a specialized rapid-sampling setup [100]. For the perturbation experiments, broth was withdrawn directly from the BioScope sample ports into the sample tubes containing the quenching solution [116]. The exact sample weights were determined by weighing each tube before and directly after sampling. Quenching and washing of the broth was done with 60% (v/v) methanol/water at -40°C and intracellular metabolite extraction was performed with the boiling ethanol method, as described in [100]. U-^{13}C-labeled cell extract was added to the pellets before extraction, as internal standard. In the samples from the reference strain, *S. cerevisiae* cell extract was used, while in samples from the mtlD strain, *P. chrysogenum* cell extract was used instead, because it contains U-^{13}C-labeled Mtl1P. Sample concentration was accomplished by evaporation under vacuum, as described in [117].

Samples for extracellular metabolite analysis - Samples for extracellular metabolite analysis were taken using the cold steel beads method as described in [116].

Intracellular metabolite analysis - The concentrations of the metabolic intermediates G6P, F6P, FBP, PEP, pyruvate, T6P, 6PG, G1P, M6P, oxoglutarate, succinate, fumarate, malate and Mtl1P, as well as the combined pools 2PG+3PG and citrate+isocitrate, were determined by ESI-LC-MS/MS [172]. Quantification was based on IDMS [117, 209]. The concentrations of NAD and NADH were determined by the spectrophotometric cycling method as described in [189].

Extracellular metabolite analysis - The concentration of glucose was measured enzymatically using Enzytec kit 1002781 (Scil Diagnostics, Germany). The concentration of acetaldehyde was measured by gas chromatography as described in [112]. The concentration of mannitol was measured by HPLC as described in [86].

Enzyme activity determination - Cell extracts were prepared according to [86], except that cell disruption was done by shaking with 0.75 g of glass beads (425-600 μm, Sigma-Aldrich) in a Qbiogene (MP Biomedicals, The Netherlands) FastPrep120 (power level 6), 4 times 20 s, at 1 min intervals, with cooling on ice in between. The activity of M1PDH was assayed according to [203]. The activity of glucose-6-phosphate dehydrogenase was also determined, according to [138]. Protein in the cell extracts was determined by the Lowry method, with BSA as standard, as described in [86].

RESULTS

M1PDH is functionally expressed at sufficiently high activity

The expression of M1PDH was confirmed by the analysis of enzymatic activity in cell extracts from the aerobic glucose-limited chemostat cultures. An *in vitro* activity of approximately 2 U/mgprotein was found in protein extracts of the mtlD strain, while no activity was measured in extracts of the reference strain (Figure 4.1). As a control, we also measured the activity of glucose-6-phosphate dehydrogenase, which was identical in extracts from both strains, at approximately 0.7 U/$mg_{protein}$.

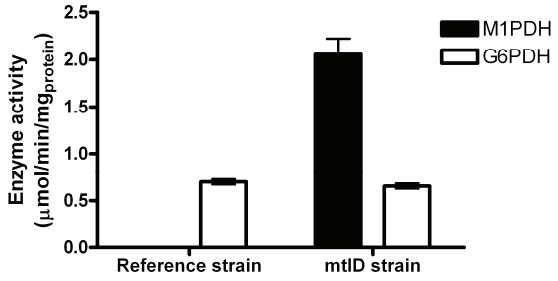


Figure 4.1: Enzyme activities of M1PDH and G6PDH measured in extracts of the reference strain and the mtlD strain from aerobic glucose-limited chemostat cultures. Values are averages of at least duplicate analysis ± standard deviation.

That M1PDH was functional *in vivo* was confirmed by the presence of an average steady-state intracellular concentration of 0.5 μmol/gDW Mtl1P and trace amounts of mannitol in extracellular samples (≈0.1 mM), in the cultures of the mtlD strain. This concentration of mannitol in the extracellular medium is comparable to the residual concentrations of other by-products such as ethanol, acetate and glycerol in aerobic glucose-limited chemostats. No

intracellular Mtl1P or extracellular mannitol was observed in the cultures of the reference strain.

The overall physiology of the chemostat cultures, in terms of biomass yield and production/consumption rates, was identical for the two strains. To check whether the introduction of the heterologous enzyme disturbed the rest of metabolism, we measured the steady-state intracellular concentrations of intermediates of central metabolism in the reference and the mtlD strain. As shown in Table 4.2, no significant differences were observed in the levels of all 17 measured metabolites, including NAD and NADH, between the two strains. This indicates that the introduction of the heterologous enzyme has no significant impact on central metabolism, at least under these growth conditions.

Table 4.2: Steady-state intracellular concentrations of measured metabolites from aerobic glucose-limited cultures at D=0.1 h^{-1}. Data are from at least 3 samplings from at least 2 independent chemostats. Each sampling comprises at least 3 replicate samples, each analyzed at least in duplicate. n.f.: not found

Intracellular concentration ($\mu mol/g_{DW}$)	Reference strain		mtlD strain	
	Average	95% confidence interval	Average	95% confidence interval
G6P	2.6	2.2 - 3	2.2	1.7 - 2.7
F6P	0.47	0.37 - 0.58	0.41	0.29 - 0.53
FBP	0.52	0.35 - 0.69	0.55	0.35 - 0.75
2PG+3PG	1.3	1.1 - 1.5	1.1	0.8 - 1.3
PEP	1	0.82 - 1.3	0.82	0.57 - 1.1
Pyruvate	0.29	0.15 - 0.44	0.43	0.26 - 0.6
T6P	0.3	0.18 - 0.42	0.26	0.12 - 0.4
6PG	0.52	0.36 - 0.68	0.5	0.31 - 0.68
G1P	0.41	0.33 - 0.49	0.34	0.24 - 0.43
M6P	1.1	0.82 - 1.4	0.84	0.53 - 1.2
Citrate	4.9	3.7 - 6.1	6	4.6 - 7.4
Oxoglutarate	0.3	0.23 - 0.37	0.31	0.23 - 0.39
Succinate	0.2	0.14 - 0.26	0.18	0.1 - 0.25
Fumarate	0.12	0.076 - 0.16	0.14	0.086 - 0.19
Malate	0.69	0.49 - 0.88	0.74	0.52 - 0.97
M1P	n.f.		0.51	0.44 - 0.58
NAD	6.8	6.5 - 7.1	6.7	6.3 - 7.1
NADH	0.96	0.59 - 1.3	0.89	0.47 - 1.3

Assuming a soluble protein content of 0.33 gprotein/gDW [138], the *in vitro* activity of 2 U/mgprotein is equivalent to an intracellular maximum activity of 11 µmol/gDW/s. As a term of comparison, that is more than 30-fold higher than the steady-state glucose uptake rate, 0.3 µmol/gDW/s. On the other hand, the flux through M1PDH estimated as the net production rate of mannitol was 0.001 µmol/gDW/s, which represents approximately 0.3% of the glucose uptake. From these values, the maximum capacity of M1PDH is over 104-fold higher than the net flux towards mannitol. Under these conditions, the reaction can be assumed to operate at near-equilibrium. Calculated from the maximum activity of the enzyme, the turn-over time of

the Mtl1P pool is less than 0.1 s. Thus, the concentration of Mtl1P can be expected to respond to sub-second changes in the concentrations of F6P, NAD and NADH.

The cytosolic free NAD/NADH ratio in yeast

As described above, the cytosolic free NAD/NADH ratio can be determined directly from the measured F6P/Mtl1P ratio and the equilibrium constant (eq.5). Using the steady-state intracellular concentrations of Mtl1P and F6P under aerobic conditions (see Table 4.2), we obtain a cytosolic free ($NAD/NADH \times 10^{pH-7}$) ratio of 101(±14) (relative to pH'=7.0).

In addition, if the cytosolic pH can be assumed or determined (and differs from pH') the NAD/NADH ratio can be calculated separately, by taking that into account. Kresnowati et al. recently estimated an intracellular average pH of about 6.5 under similar growth conditions (D=0,05 h-1, Cx=14.5 gDW/l) [95]. Preliminary results using a new fluorescence-based technique indicate a cytosolic pH, under the same growth conditions used in this study, of about 6.7 (R. Orij and J. Postmus, University of Amsterdam, personal communication). Assuming a cytosolic pH value of 6.5, the cytosolic free NAD/NADH ratio is 320(±45), while with a pH of 6.7, it is 202(±28). These values are more than 10-fold higher than the total NAD/NADH ratio of 7.5(±2.5) calculated from the measured whole-cell total concentrations of NAD and NADH (see Table 4.2).

Table 4.3: Influence of the NAD/NADH ratio on the thermodynamic feasibility of glycolysis. Concentrations were converted to mM using the factor 2,38 mlcell/gDW [170]. The cytosolic free NAD/NADH was calculated from the steady-state ($NAD/NADH \times 10^{pH-7}$) ratio of 101, assuming the cytosolic pH is between 6.5 and 7.0 (see Results). Gibbs energies of formation of the metabolites at pH' 7.0 and 6.5, T=25°C and I=0.25 M were obtained from [3]. The Gibbs energy of reaction was calculated for the overall reaction: FBP + 2 NAD + 2 ADP + 2 Pi → 2 3PG + 2 ATP + 2 NADH. A reaction is feasible if $\Delta rG' \leq 0$.

Strain and conditions	CEN.PK 113-7A:mtlD, aerobic, glucose-limited, $D=0.1h^{-1}$	
	with whole-cell total NAD/NADH	with cytosolic free NAD/NADH
FBP (mM)	0.23	
3PG (mM)	0.41 [a]	
ATP (mM)	3.06 [b]	
ADP (mM)	0.72 [b]	
Pi (mM)	43 [b]	
whole-cell total NAD/NADH	7.5	
cytosolic free NAD/NADH (assuming cytosolic pH 7.0 / 6.5)		101 / 320
$\Delta_r G'$ (kJ/mol) (at pH' 7.0 / 6.5)	+12 / +17	-1.3 / -1.4

a: calculated from the concentration of 2PG+3PG assuming equilibrium of phosphoglycerate mutase, with $\Delta rG'^{\circ} = 5.9$ kJ/mol [3]. b: taken from [210].

To test whether the values obtained for the cytosolic free NAD/NADH ratio are realistic, we performed a thermodynamic analysis of central glycolysis (Table 4.3). We focused particularly

on the influence of the NAD/NADH ratio on the overall ΔrG' of the steps between FBP and 3PG. If the whole-cell total concentrations of NAD and NADH are used, the calculated ΔrG' of the overall reaction is highly positive. Given the known direction of the glycolytic flux, this would contradict the 2nd law of thermodynamics. Instead, if the cytosolic free ($NAD/NADH \times 10^{pH-7}$) determined above is used the calculated ΔrG' is slightly negative, in agreement with the direction of flux in glycolysis. In fact, the value of ΔrG' is close to zero, suggesting that at steady-state this set of reactions operates close to equilibrium. Thus, the values found for the cytosolic free NAD/NADH ratio are a more realistic description of the *in vivo* redox state of the NAD-NADH couple than the whole-cell total concentrations of NAD and NADH. Note also that because the ($NAD/NADH \times 10^{pH-pH'}$) ratio already includes the pH the thermodynamic analysis of networks containing redox reactions becomes much less dependent of information on the pH (Table 4.3).

The cytosolic free ($NAD/NADH \times 10^{pH-pH'}$) ratio responds rapidly to changes in electron-donor and electron-acceptor

If the activity of M1PDH is high enough, it should also be possible to use this method under highly dynamic conditions. We therefore applied this system to study the influence of external perturbations on the redox state of the cytosolic free NAD/NADH. We focused in particular on the timescale <2 min, where the initial metabolic dynamics in response to external stimuli are known to occur [170, 189]. First, a sudden increase in electron-donor was accomplished by applying a 2.3 mM glucose pulse to steady-state glucose-limited cells ([glucose]residual≈0.13 mM). Second, a simultaneous increase in electron-donor and electron-acceptor was carried out by applying a 2.3 mM glucose + 2.3 mM acetaldehyde pulse. These two perturbations are thought to affect the NAD/NADH ratio in different ways [112]. The glucose pulse is thought to lead to an accumulation of NADH (lower NAD/NADH ratio). The addition of acetaldehyde, which can enter the cells and be reduced to ethanol (with NADH as coenzyme), should cause the NADH to be re-oxidized faster, or even depleted (smaller drop or even increase in NAD/NADH ratio). The results of these experiments are shown in Figure 4.2.

The dynamic responses of the cytosolic free ($NAD/NADH \times 10^{pH-7}$) ratio (relative to pH'=7.0) were calculated from the measured profiles of F6P and Mtl1P. The profiles of the total NAD/NADH ratio were calculated from the measured whole-cell concentrations of NAD and NADH. We found that the cytosolic free ($NAD/NADH \times 10^{pH-pH'}$) ratio reacted rapidly to the perturbations. The glucose pulse led to a 4-fold increase in glucose uptake, from 0.3 to 1.2 μmol/gDW/s. The cytosolic free ($NAD/NADH \times 10^{pH-pH'}$) ratio decrease by 2-fold in less than 4 s. The glucose + acetaldehyde pulse also increased the glucose uptake, to 1.3 μmol/gDW/s, and led to an acetaldehyde uptake rate of approximately 3.4 μmol/gDW/s. In this case, the cytosolic free ($NAD/NADH \times 10^{pH-pH'}$) ratio increased more than 3-fold within 14 s. These observations are in good agreement with the expectation that a sudden increase in the uptake of glucose (that is, in the supply of electrons) would lead to a lower redox state in the cytosol, while the addition of an extra redox sink such as acetaldehyde would minimize that effect or even reverse it.

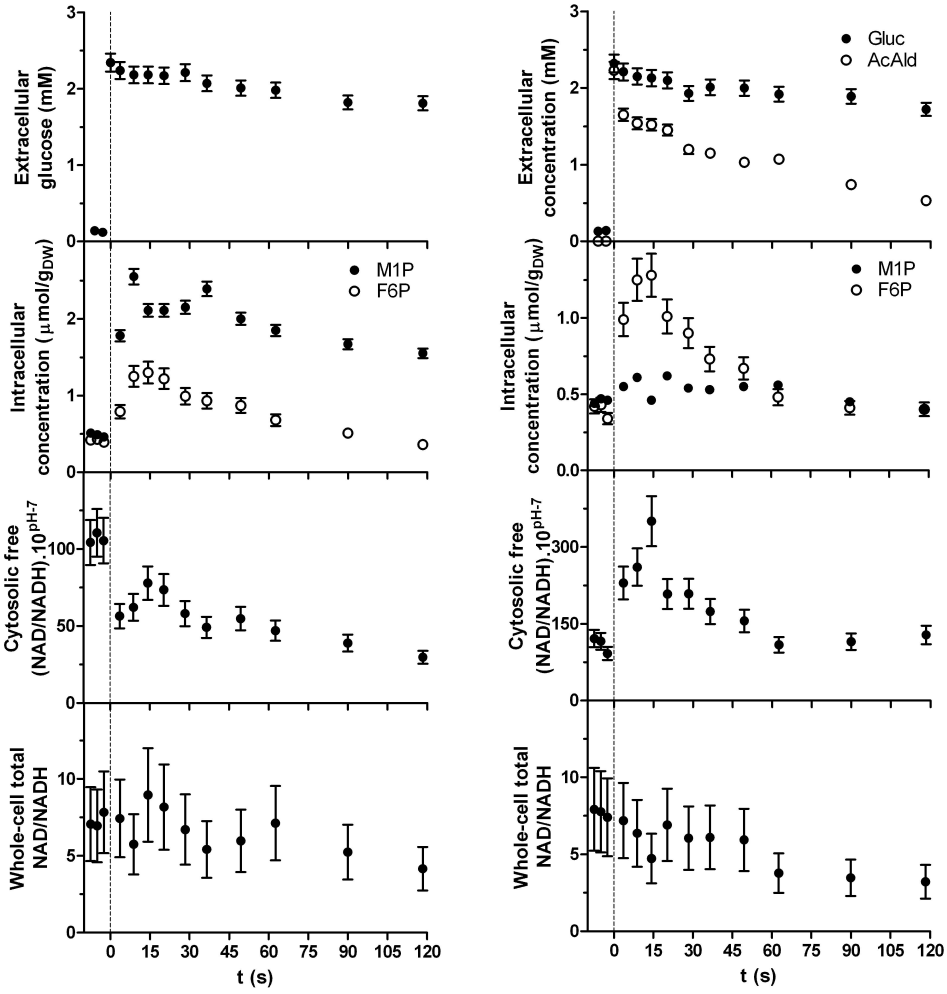


Figure 4.2: Perturbation response profiles after a 2.3 mM glucose pulse (left panel) and a 2.3 mM glucose + 2.3 mM acetaldehyde pulse (right panel) applied to aerobic glucose-limited cultures of the mtlD strain. The cytosolic free (NAD/NADH).$10^{pH-7.0}$ ratio (relative to pH'=7.0) was determined from the measured concentrations of F6P and Mtl1P and the total NAD/NADH was calculated from whole-cell total concentrations of NAD and NADH. Data points before t=0 s represent the steady-state before the perturbation and data points at t=0 s represent the calculated concentration(s) at the moment of the pulse. Dashed vertical lines indicate t=0 s. Values are averages of at least duplicate analysis ± standard deviation.

Note that the profile of F6P is almost not affected by the addition of acetaldehyde, as reported previously [112]. It is the concentration of Mtl1P that responds differently to the two perturbations, demonstrating that it is influenced by the effect of acetaldehyde on the redox state of the cytosolic NAD-NADH couple. In addition, the concentration of Mtl1P can change very rapidly. Immediately after the glucose pulse, it increased at a rate of approximately 0.4 µmol/gDW/s. That is the second highest rate of change observed in any of the 18 metabolite pools measured in our experiments, after G6P. It also represents an immediate 400-fold

increase from the net rate at steady-state, as calculated from the formation of mannitol. That shows that there is ample extra capacity of M1PDH and indicates that the reaction does operate near equilibrium.

Contrary to the cytosolic free ($NAD/NADH \times 10^{pH-pH'}$) ratio, the whole-cell total NAD/NADH ratio did not change significantly. That is in agreement with previous literature where NAD and NADH were measured after glucose pulses [170, 189]. An apparent decrease after the glucose+acetaldehyde pulse actually seems to contradict the expected effect of the acetaldehyde addition. Both observations indicate that the NAD/NADH ratio from total whole-cell concentrations of NAD and NADH does not adequately reflect the redox state of the cytosolic NAD-NADH couple.

DISCUSSION

The investigation of metabolic networks under *in vivo* conditions relies on the ability to obtain reliable data on the state of the system of interest. For the modeling of *in vivo* enzyme kinetics, in particular, we need accurate concentrations of the substrates and effectors involved in the reactions or pathways to be studied. The fact that based on the published data even a central, well-known pathway such as glycolysis is found to be thermodynamically infeasible (Table 4.1) is a clear sign that the data currently available is not reliable enough, or at least that some important information is missing. In their analysis of the problem, Maskow and von Stockar [118] dismissed the importance of the metabolite concentrations, NAD and NADH in particular, because the pathway would only become thermodynamically feasible given intracellular concentrations that they considered "unrealistic". Instead, they proposed the uncertainty of the standard Gibbs energies of reaction and neglecting the activity coefficients as the most important sources of error. Certainly, reliable standard Gibbs energies of reaction, which are obtained from the equilibrium constants, are essential. But equilibrium constants can be determined *in vitro* under well-controlled conditions. On the contrary, obtaining (meaningful) intracellular concentrations is still a challenging task.

The works of the Krebs group and others in the 1960s-70s showed that the NAD/NADH ratio calculated from whole-cell concentrations of NAD and NADH does not reflect the *in vivo* redox state of the NAD-NADH couple in different cell compartments [26, 164, 182]. The free NAD/NADH in each cell compartment, as determined from suitable sensor reactions, seems to be a much more reliable measurement. Reported values for the cytosolic free NAD/NADH can be up to 100-fold higher than the total NAD/NADH [199]. Such differences are sufficient to close the $\Delta rG'$ gap found in the thermodynamics of glycolysis, indicating that the NAD/NADH ratio could be the source of the error.

For M1PDH to be used as sensor of the cytosolic free NAD/NADH, is must satisfy all the necessary criteria mentioned in the Introduction. Functional expression of mtlD in our strain was confirmed by the measurement of *in vitro* activity of the enzyme in protein extracts (Figure 4.1) and the presence of an intracellular pool of 0.5 μmol/gDW of Mtl1P. The estimated *in vivo* maximum activity was about 104 higher than the net flux. This is because the enzyme is being

over-expressed but in the absence of a mannitol 1-phosphatase it catalyzes essentially a dead-end reaction. Furthermore, the high rate of change in the Mtl1P concentration under transient conditions (Figure 4.2) demonstrates that there is ample extra enzyme capacity. So we can conclude that the enzyme activity is high enough to establish near-equilibrium *in vivo*. Furthermore, the enzyme used, M1PDH from *E. coli*, is strictly specific for its substrates and coenzyme [127, 168, 205]. In the absence of specific targeting, the enzyme should be expressed only in the cytosol. In addition, the reactants F6P and Mtl1P can be reasonably assumed to be both strictly cytosolic, so the reaction can only take place in the cytosol. Thus, the system described here fully satisfies all necessary criteria to constitute a reliable sensor of the cytosolic free NAD/NADH ratio.

Since a new reaction had to be introduced, it was preferable if it had as little impact on the rest of metabolism as possible. *S. cerevisiae* does not express a native M1PDH. Accordingly, we observed no mannitol or Mtl1P in samples from the reference strain (Table 4.2). Previous studies showed that expression of mtlD from *E. coli* in wild-type *S. cerevisiae* has no observable phenotypic consequences, in terms of morphology, growth characteristics or glycerol production [36]. If an unimpaired glycerol metabolism is present, it results only in the production of minor amounts of mannitol, presumably by un-specific de-phosphorylation of Mtl1P [41]. In Lactococcus lactis, the introduction of a mannitol 1-phosphatase proved to be the key factor for effective production of mannitol [203]. Without a specific mannitol 1-phosphatase, M1PDH should operate virtually as a dead-end pathway, thus minimizing any impact on the rest of metabolism. Accordingly, we observed no physiological difference between the mtlD strain and the reference in aerobic glucose-limited chemostats. Only residual amounts of mannitol were found in the medium, as expected. Furthermore, we also compared the intracellular levels of intermediates of central metabolism. The concentrations of all measured metabolites, including F6P, NAD and NADH, were unchanged, compared to the reference strain (Table 6.2). From the available literature and our own results, we can conclude that the introduction of M1PDH has no measurable impact on the rest of metabolism, at least under these growth conditions. In our view, this is sufficient to assume that observations in the constructed strain should also be valid for the reference strain.

Under aerobic glucose-limited steady-state conditions at μ=0.1 h-1 we obtained a cytosolic free NAD/NADH ratio between 101($\pm$14) and 320($\pm$45), assuming the cytosolic pH is between 7.0 and 6.5, respectively. To our knowledge, this has never been reported before. Under the same conditions, we obtained a total NAD/NADH ratio of 7.5($\pm$2.5) from the whole-cell concentrations of NAD and NADH. This value is in the range reported for *S. cerevisiae* under similar growth conditions [170, 189]. The cytosolic free NAD/NADH is thus more than 10-fold higher than the total NAD/NADH. Such a large difference was also observed in rat liver cells, where the cytosolic free NAD/NADH ratio was up to 100-fold higher than determined by total NAD and NADH [199]. The higher cytosolic free NAD/NADH is thought to be due to uneven distribution of NAD and NADH between compartments and a large fraction of cytosolic NADH being protein-bound [26]. These effects cannot be observed by direct measurement of total concentration of NAD and NADH.

Thermodynamic analysis of the reactions of central glycolysis with the whole-cell total NAD/NADH ratio shows that the data contradicts the second law of thermodynamics (Table

4.3), like the datasets from published literature (). On the contrary, the analysis performed using the cytosolic free NAD/NADH ratio shows that the dataset becomes thermodynamically consistent. This difference in outcomes demonstrates that the values found for the cytosolic free NAD/NADH ratio are a much more realistic description of the redox state of the cytosolic NAD-NADH couple.

It should be noted that in these thermodynamic calculations we have used the whole-cell concentrations of ATP, ADP and phosphate, which are also present in different compartments and bound by abundant enzymes. In rat hepatocytes the influence of compartmentation and protein-binding on the ATP/(ADP.Pi) ratio was found to be not very significant [164]. Under normal conditions, the calculated cytosolic free ATP/(ADP.Pi) ratio was within ±30% of the total ATP/(ADP.Pi) ratio. That is equivalent to a possible variation of ±1.3 kJ/mol in the $\Delta rG'$ calculations above which, although not negligible, is not sufficient to change their outcomes.

Short-term perturbation experiments are particularly attractive for *in vivo* kinetic parameter estimation because they are information-rich [97, 125, 149]. However, it is essential that the relevant system dynamics be accurately captured in the measured dynamic metabolite data. In previous literature reporting dynamics of intracellular metabolites after glucose pulses to glucose-limited yeast cultures, the whole-cell concentrations of NAD and NADH were found to remain more or less constant [170, 189]. This was considered somewhat unexpected, since the sudden increase in glucose uptake and the triggering of ethanol production are compatible with a transient decrease in NAD/NADH ratio. But in agreement with these findings, we also observed no significant changes in total NAD/NADH ratio in our glucose pulse experiment (Figure 4.2). Furthermore, in the glucose+acetaldehyde pulse experiment the total NAD/NADH ratio actually seemed to decrease slowly, in contradiction with the expected role of acetaldehyde as a redox sink. On the contrary, the cytosolic free ($NAD/NADH \times 10^{pH-pH'}$) ratio changed rapidly upon both perturbations. In addition, it changed upwards or downwards in agreement with expectations. Upon the glucose pulse it decreased, indicating the cytosol was more reduced, and upon the glucose+acetaldehyde pulse it increased, indicating the cytosol was more oxidized. Note that the profile of F6P is not very affected by the addition of acetaldehyde, as reported before [112]. It is the profile of Mtl1P that differs between the experiments. Since the reaction operates near equilibrium, the concentration of Mtl1P depends only on the concentration of F6P and the ($NAD/NADH \times 10^{pH-pH'}$) ratio. If the profile of F6P is similar in the two experiments, the difference in Mtl1P profile must be the reflection of changes in the redox couple. Thus, the F6P/Mtl1P ratio is a reliable indicator of the cytosolic free ($NAD/NADH \times 10^{pH-pH'}$) ratio also in highly dynamic conditions.

Unfortunately, it is not yet possible to calculate the NAD/NADH ratio separately for the pulse response data. This would require reliable information on the short-term dynamics of the cytosolic pH, which is neither available from the literature nor currently possible to determine. Previous determinations of the average cellular pH of yeast have suggested a fast acidification occurs within a few seconds after a glucose pulse, followed by slow alkalinization [95], although a plausible mechanistic explanation for this observation has yet to be found. Fluorescence-based methods to measure compartment-specific pH may provide the necessary information in the near future (R. Orij and J. Postmus, University of Amsterdam, personal communication).

Without such information, it cannot be excluded that changes in cytosolic pH may be partly responsible for the dynamic response observed in the ($NAD/NADHx10^{pH-pH'}$) ratio.

The fact that we detect substantial dynamics in cytosolic free ($NAD/NADHx10^{pH-pH'}$) ratio while the total NAD/NADH ratio does not change much may appear contradictory, but these observations are not irreconcilable. Since a major fraction of cytosolic NADH ratio is likely protein-bound [26], the reduction of even a small amount of NAD is sufficient to change the free concentration of NADH significantly. However, that would change the total concentration of NADH by only a small amount. Thus, the cytosolic free NAD/NADH ratio can change substantially without large changes in the total NAD and NADH concentrations. Zhang et al. have proposed that this property of the cytosolic NAD-NADH couple makes it a sensitive redox sensor for transcriptional regulation [212]. Due to the number of reactions involving NAD as coenzyme, the cytosolic NAD-NADH couple is also likely to play an important role in regulation at the metabolic level. Since it is the free coenzyme that is kinetically and thermodynamically relevant, the method presented here provides new, more reliable data for quantitative analysis of metabolic networks in *S. cerevisiae*.

CONCLUSIONS

To our knowledge, this is the first report of the expression of a heterologous enzyme for the purpose of determining the free NAD/NADH ratio in a specific compartment from the concentration ratio of its reactants. The method we describe can be used at steady-state as well as in highly dynamic conditions and constitutes the most reliable estimate of the redox state of the NAD-NADH couple in the cytosol of *S. cerevisiae*. We think it will provide new insight on the role of the NAD/NADH ratio in regulation and signaling mechanisms and invaluable data for *in vivo* kinetic modeling and biothermodynamics.

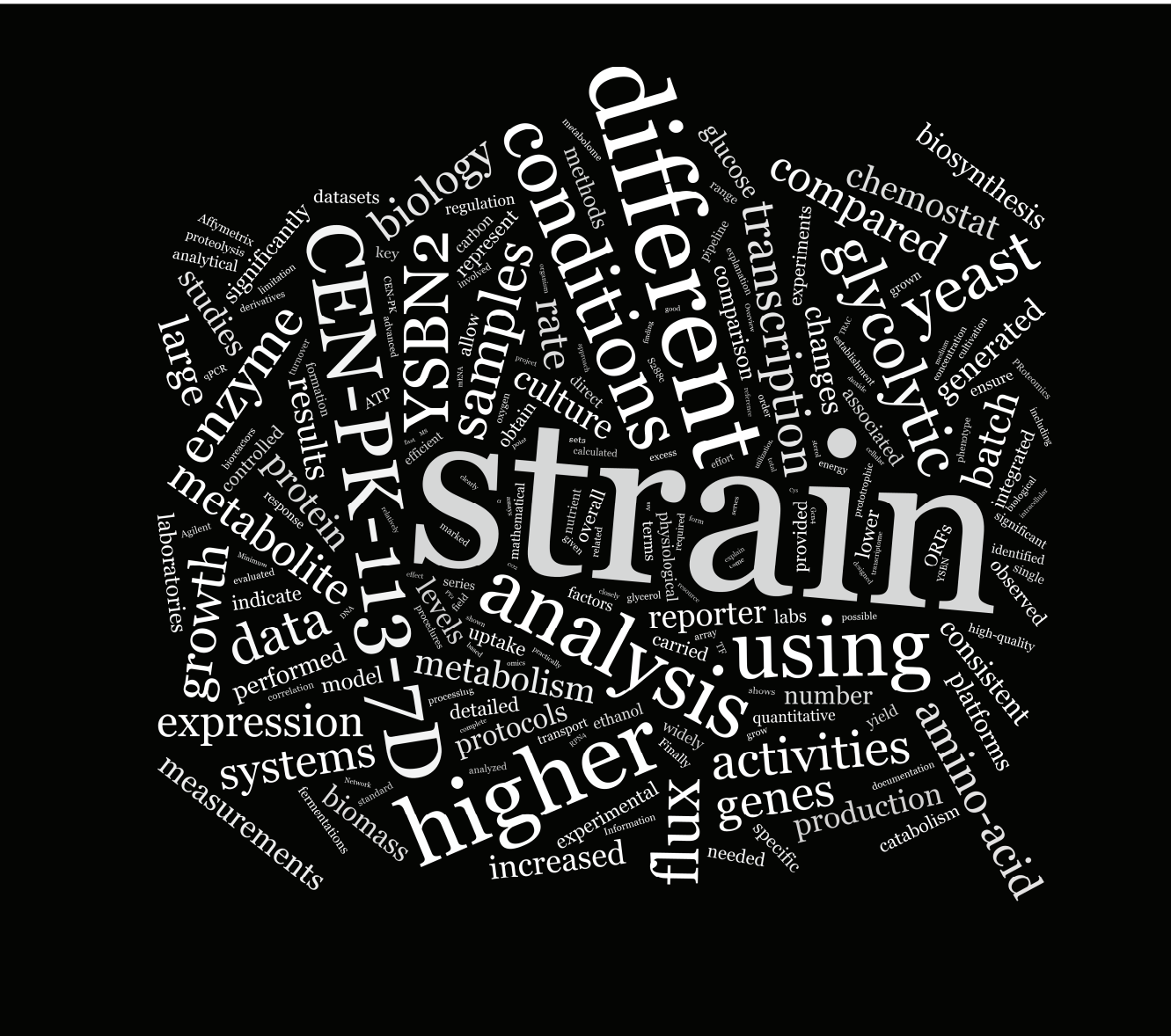
ACKNOWLEDGEMENTS

We thank Roeland Costenoble for promptly providing the pYX222:mtlD plasmid and Ron Winkler for competently constructing the mtlD strain. Hans van Dijken and Jack Pronk are acknowledged for their valuable advice in the screening for suitable sensor reactions. We also thank Zheng Zhao for making P. chrysogenum U-^{13}C-labeled cell extract readily available and Marijke Luttik for assistance with the enzyme activity assays. Finally, we are grateful to Cor Ras and Johan Knoll for the outstanding LC-MS work and Peter Verheijen for help with statistical data analysis. This work was funded by SenterNovem through the IOP Genomics initiative (project IGE3006A).

Chapter 5

Can we put all our data together?

Lessons from an integrated multi-laboratory comparison of platforms for experimental Systems Biology in yeast



ABSTRACT*

Despite much progress in the field of systems biology, this field is often held back by difficulties in obtaining comprehensive, high-quality, quantitative datasets. The yeast *Saccharomyces cerevisiae* serves as an excellent model organism in the field of systems biology. Besides its important role as a model organism it also serves as an industrial work horse in the production of fuels, chemicals, food ingredients and pharmaceuticals. Here we describe the results of an inter-laboratory effort to generate high-quality quantitative data for a very large number of cellular components in yeast. We ensured the high-quality of experimental data by evaluating a wide range of sampling and measurement techniques. The data were generated for two different laboratory yeast strains, each growing under two different growth conditions and this large dataset was used to perform an integrated analysis that provided new insights into yeast metabolism.

* Submitted as: Experimental systems biology: Lessons from an integrated, multi-laboratory study in yeast.
This work was carried out under the auspices of the Yeast System Biology Network Coordination Action (ysbn.eu) funded through the EC's FP6 and headed by Prof. Jens Nielsen at TU Chalmers, Sweden

INTRODUCTION

There are many definitions and interpretations of systems biology, but most involve mathematical modelling, high-throughput (or omics) analysis, mapping of interactions between cellular components, and quantification of dynamic responses in living cells[9, 22, 23, 83, 93]. In most cases the objective of systems biology is to obtain a quantitative description of the biological system under study, and this quantitative description is ideally in the form of a mathematical model that can be used to simulate the operations of the biological system. Even though some mathematical modelling concepts rely only on limited datasets, (e.g. flux balance analysis) most systems biology efforts will require large sets of high-quality experimental data that enable, for example, to discriminate between different model structures. Generation of such data is therefore the core of many studies that use the systems biology approach. However, the infrastructure and know-how needed to generate the large number of different data required for advanced systems biology studies (e.g. transcriptomics, proteomics, metabolomics) is normally beyond the capabilities of a single lab. There is therefore a trend towards multi-lab collaboration projects and the establishment of curated databases that contain high-quality datasets[146]. In order to ensure proper documentation of experiments, some effort has been directed also at establishing protocol formats, such as MIAME (Minimum Information About a Microarray Experiment) for DNA array experiments[21 7], MIAPE (Minimum Information About a Proteomics Experiment) and PRIDE (PRoteomics IDEntification) for proteome analysis[110, 165], protocols for microbial metabolome analysis[175], and even protocols for documentation of mathematical models such as MIRIAM[103] (Minimum Information Requested In the Annotation of biochemical Models). Even though these protocol formats aim to ensure proper documentation of the actual experiments, there is still a need for consolidation of applied experimental conditions and procedures, in order to allow the generation of increasingly large, coherent datasets for the same organism or strain, that will eventually represent a rich resource for advanced mathematical modelling and contribute to our understanding of the living cell.

The Yeast Systems Biology Network (YSBN) therefore undertook a major effort on consolidating and comparing experimental conditions, procedures and protocols applied for yeast systems biology in 10 different European laboratories, and at the same time performed a comparative analysis of different quantitative analytical methods. This has resulted in establishment of a well documented experimental "systems biology pipeline" that is illustrated in Figure 5.1 The "pipeline" allows for the comparison of different yeast strains or the comparison of a single yeast strain grown under different conditions. Here we evaluated the "pipeline" by comparing two different yeast strains grown at two different conditions in bioreactors, namely a traditional batch culture (nutrient excess) and a glucose-limited chemostat culture (specific growth rate controlled by the rate of supply of the limiting nutrient, glucose). The generated data will represent valuable reference data for the two strains and two standard conditions and hence advance the field of yeast systems biology.

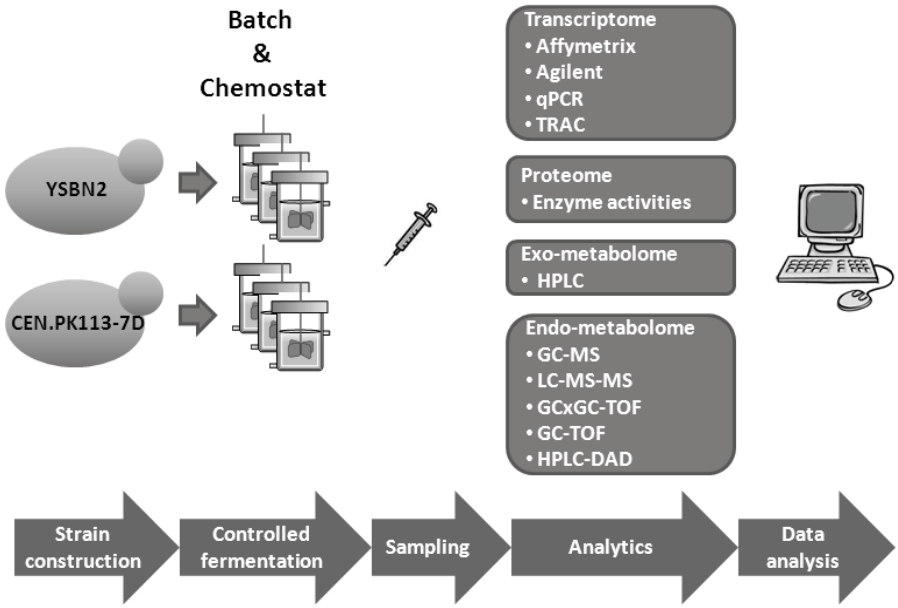


Figure 5.1: Overview of the "systems biology pipeline" used in this study. Two different haploid yeast strains are grown at two different conditions and samples are taken for several kinds of analysis. In the experiment described here, two yeast strains were used: YSBN2, that is closely related to the originally sequenced yeast strain S288c (for strain construction see Supplementary 1) and CEN.PK113-7D, that is a widely used yeast strain for physiological studies and industrial applications. Each strain was grown in bioreactors, both under substrate excess growth conditions (batch) and also under glucose-limited conditions (chemostat) (see Supplementary 2 for details). These fermentations were carried out in biological triplicates at a single location (TU Delft) and by using highly controlled bioreactors it was possible to obtain reproducible measurements of key physiological parameters, such as specific growth rate, nutrient uptake rates and product formation rates (see Supplementary 3). A fast sampling protocol was carefully designed (see Supplementary 4) taking into account the large amounts of samples needed as well as the turnover of the different types of molecules to be analyzed. After sampling and conditioning the different samples were shipped to the different laboratories involved in this project for analysis of the transcriptome, enzme activities, the metabolome and the activity key enzymes. Finally the results were collected and integrated, resulting in a very thorough phenotypic characterization of the two strains.

RESULTS & DISCUSSION

The first step in the establishment of our experimental systems biology pipeline was to find appropriate yeast strains that would be of interest. In the yeast community a range of different yeast strains are being employed, with the strain series BY, W303 and CEN.PK being the most frequently used[176]. The BY strain series is a direct derivative of the originally sequenced strain S288c, for which there is a complete gene knock-out collection. The CEN.PK strain series is used widely in physiological studies and, thanks to its rapid growth, it is also often used for metabolic engineering studies[176]. Physiological studies are generally best performed with prototrophic strains, whereas the BY series strains carry a number of auxotrophies that may cause problems for quantitative studies of the cellular physiology. There was hence a need to generate prototrophic strains in the S288c background that nonetheless carried some genetic markers that would permit checks against contamination, in large-scale or long-term cultures, and facilitate subsequent genetic crosses (for instance of evolved derivatives). For this reason,

two diploid strains, YSBN1 and YSBN2, were generated from FY1 and FY2, two uracil auxotrophic strains that are direct derivatives of S288c[202]. FY2 is the strain from which the BY strain series[18] and, hence, the strain from which the complete knockout collection[63] was derived. YSBN1 and YSBN2 are prototrophic strains that carry drug-resistance markers inserted into their genomes at a phenotypically neutral site[6, 130] (Supplementary 1).

In order to evaluate the newly constructed strain we performed a detailed comparative analysis of the YSBN2 strain with the widely used strain CEN.PK113-7D. Each strain was grown in batch cultures, which were sampled in the mid-exponential growth phase (on glucose). After the diauxic shift and the ethanol growth phase the cultures were used to initiate chemostat cultivations, which were sampled once steady state conditions were achieved. The fermentations were carried out in triplicate with each strain and each condition, resulting in a total of 12 samplings. The detailed cultivation protocols are given as Supplementary 2. The experiments were conducted in well controlled bioreactors in a single laboratory, which ensured a very high degree of reproducibility (Supplementary 3), and samples were then shipped to different laboratories for analysis (see Table 3 in Supplementary 3). From the initial analysis it is interesting to observe that the maximum specific growth rate of the CEN.PK strain is significantly higher than that of the YSBN strain, by approximately 25%, whereas its biomass yield under carbon-limited conditions (chemostat) is significantly lower, by approximately 10%. Thus, one could speculate whether CEN.PK's ability to grow faster when resources are abundant has come at the expense of efficiency in carbon and energy utilization under nutrient limitation.

Our analysis involved sampling for analysis of: 1) mRNA; using DNA arrays (Affymetrix and Agilent), qPCR and TRAC (TRanscript analysis with Affinity Capture)[143]; 2) enzyme activities; using optimized and "*in vivo*-like" assays; and 3) endometabolome, using several analytical platforms including LC-MS/MS, GC-TOF, GCxGC-TOF, NMR, HPLC-DAD and enzymatic analysis. Table 1 in Supplementary 4 gives an overview of the sampling procedure, which was designed taking into account the large number of samples needed and the fast turnover of some the molecules to be analyzed (e.g. intracellular metabolites). Fast sampling and parallel sample processing required a team of 5 people, of which 3 carried out the same key steps in all 12 samplings. Videos demonstrating the sampling are available at www.sysbio.se/supp. Detailed protocols of the sampling and processing methods are given in Supplementary 4.

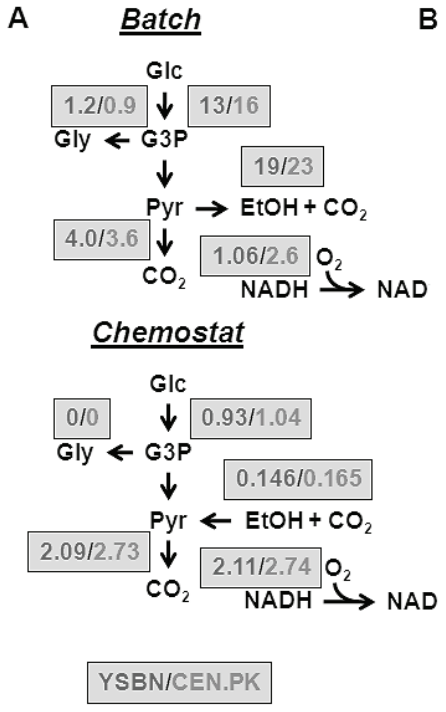
Transcripts were measured using a range of different methods. Genome wide analysis was performed using both the Affymetrix and the Agilent platforms (Supplementary 5). qPCR was performed in two different laboratories and was used to quantify the expression of a number of selected genes (Supplementary 6 and Supplementary 7). TRAC analysis was also performed, as it allows for multiplex detection of mRNA targets simultaneously from a large number of samples (Supplementary 8). Comparison of mRNA levels determined in four different labs using the four different analytical methods shows a good overall consistency (Fig. 5 in Supplementary 5). qPCR analysis performed in one of the labs was used to check the consistency of 33 genes that were found to have significantly changed expression based on an ANOVA analysis of the Affymetrix data. This analysis shows very clearly that the qPCR analysis can confirm the direction of expression for all these genes (Table 7 in Supplementary 7). The TRAC analysis also showed a fairly good consistency with the DNA array data, but

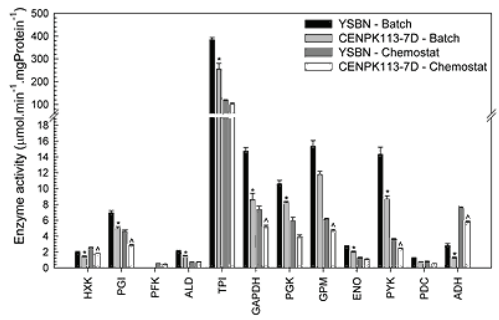
about 15% of the analyzed transcripts had a very poor correlation, based on analysis of the four different samples (2 strains x 2 conditions), and another 15% had a Pearson correlation coefficient lower than 0.85 (Table 2 in Supplementary 8). We did not find a linear correlation between the Affymetrix and the Agilent data (Fig. 1 and Fig. 5, Supplementary 5), but this can be explained by the use of a different dynamic scanning range by the Agilent scanner. However, the significant genes identified by ANOVA analysis using the two platforms were fairly consistent, i.e. out of a total of 410 transcripts found to be significantly changed in response to growth conditions (including both strains and both array platforms) 241 transcripts were found consistently by both platforms (Fig. 2 in Supplementary 5). Thus, we conclude that both platforms are equally strong, but in light of the large datasets available already with the Affymetrix we recommend the use of this platform.

We also performed activity measurement of key glycolytic enzymes in cell extracts. Measurements carried out in two different laboratories yield a good overall consistency in the data (Fig. 1 and 2, in Supplementary 9). In connection with the experiment we also evaluated a new approach to quantify enzyme activities using assay conditions designed to better represent *in vivo* conditions, as opposed to the typical approach of using optimal conditions for each enzyme (Fig. 3 in Supplementary 9). We found that with the "*in vivo*-like" assays the range of enzyme activities was in the same order of magnitude as the glycolytic flux, which was not observed using traditional analysis of glycolytic enzyme activities.

Metabolome samples were analyzed independently by seven participating labs according to their own standard operating procedures (see Supplementary 9). A total of four variations in extraction protocol, five analytical platforms (LC-MS/MS, GC-TOF, GCxGC-TOF, enzymatic, HPLC-DAD), and three internal standard strategies were used in different combinations. The interlab comparison shows that absolute concentration estimates from different labs can vary by up to 3-fold, even for identical sample processing. Interestingly, relevant metabolite ratios (e.g. mass action ratios) based on measurements from different labs were comparable. Furthermore, when comparing only the relative concentration differences between growth conditions or strains, all labs deliver a consistent picture. Since this redundancy was obtained with heterogeneous preparation protocols and orthogonal analytical methods we conclude that, at the current state-of-the-art, ratio-metric measurements still have much higher confidence than absolute estimates.

Integrated data analysis started with looking at the exo-metabolome data, which provided information about the "gross phenotype". Through measurements of all key nutrients and metabolites being secreted or released into the medium (including analysis of the gas-phase) it was possible to obtain very precise measurements of the overall metabolic fluxes and these are summarized in Figure 5.2 (panel A). Under substrate-excess growth conditions (batch cultivation) it was observed that the CEN.PK113-7D strain exhibited a >20% higher glucose uptake rate, which was accompanied by a higher flux towards ethanol. The specific carbon dioxide production not associated with ethanol production was the same in the two strains, despite the fact that the specific oxygen uptake rate was more than 2-fold higher in the CEN.PK113-7D strain when compared to the YSBN2 strain. This suggest that the higher oxygen uptake rate allowed for a reduced glycerol production by the CEN.PK113-7D strain, as





C

$\log2\left(\frac{\text{batch}}{\text{chemostat}}\right)$

Metabolome data	CEN.PK	YSBN2
Glucose	6.8	8.0
Glucose 6-P	0.6	1.0
Fructose 6-P	-0.2	0.4
Fructose 1,6-bisP	4.9	5.5
Dihydroxyacetone-P	2.8	3.5
Glycerol 3-P	2.7	3.6
2,3-Bis-P-glycerate	0.4	-0.3
3-P-glycerate	-1.9	-3.1
2/3-P-glycerate	-1.8	-2.6
Phosphoenolpyruvate	-2.4	-2.7
Pyruvate	4.2	4.5
Lactate	2.5	3.8

Figure 5.2: Overall fluxes in the strains CEN.PK113-7D and YSBN2 in batch and chemostat cultivations. **(A)** From measurements of the concentrations of biomass, glucose, ethanol and glycerol in the medium and of carbon dioxide and oxygen in the exhaust gas, the overall exchange rates can be calculated (all data are given in Supplementary 2). From the overall fluxes it is observed that the CEN.PK113-7D strain has a higher glucose uptake rate, a higher flux towards ethanol and a higher oxygen uptake rate, whereas the carbon dioxide production rate is about the same for the two strains (the flux towards CO_2 is corrected for the amount of CO_2 formed in relation to ethanol production). However, YSBN2 has a higher glycerol production rate in the batch fermentation. In the chemostat culture the biomass yield is higher for the YSBN2 strain and this is accompanied by higher specific fluxes in the CEN.PK113-7D strain. (B) The higher glycolytic flux in the CEN.PK113-7D strain at substrate excess growth conditions (batch) is not associated with increased activities of glycolytic enzymes. For practically all of the glycolytic enzymes there was a higher activity in the YSBN2 strain compared with the CEN.PK113-7D strain (results were confirmed by measurements in two independent laboratories, see Supplementary 8). In the comparison between growth conditions, the higher glycolytic flux in batch culture compared with the chemostat is associated with higher enzyme activities for practically all glycolytic enzymes in both strains. However, this increase of more than 10-fold in glycolytic flux is accomplished with relatively modest changes in enzyme activities, pointing to a substantial element of metabolic regulation. (C) This is supported by measurement of glycolytic metabolites, where it is found that for both strains there is an increase in most of the glycolytic metabolites in the batch culture compared with the chemostat (see Supplementary 11 for more details). Thus, the dramatic increase in glycolytic flux in yeast when there is a shift from glucose limitation to glucose excess is due partly to an increase in the activity of the glycolytic enzymes, but also due to a kinetic effect resulting in increased levels of most glycolytic metabolites. This finding is consistent with earlier more detailed studies on glycolysis[43, 139, 173, 178]. The metabolite data, however, do not provide any insight into the differences in fluxes between the two strains, as there is very small differences in metabolite levels between the different strains, and most differences indicate slightly higher metabolite concentrations in the YSBN2 strain, which has lower fluxes.

overflow towards glycerol production under fermentative growth conditions serves as an alternative NADH sink. The higher glycolytic flux in the CEN.PK113-7D strain was not associated with a higher activity of glycolytic enzymes or large differences in the levels of glycolytic intermediates (Figure 5.2, panels B and C). In contrast, in both strains we found higher enzyme activities under batch conditions than in the chemostat. However, since the enzyme activity level was only about 50% lower in the chemostat compared with the batch

cultures, it is clear that enzyme activities can not alone describe the almost 10-fold lower glycolytic flux in the chemostat. This indicates that the adjustment of the glycolytic flux to fermentative conditions is to a large extent determined by changes in the levels of metabolic intermediates and effectors, i.e. flux control is primarily at the metabolic level. This observation is compatible with a general pattern of metabolite-dominated regulation in the central metabolism of yeast[43, 139, 173, 178]. Finally, the higher activity of glycolytic enzymes in the batch cultures is associated with increased expression of several glycolytic genes (*PFK1*, *TPI1*, *ENO2*, *TDH2*, *TDH3*, *PYK1*, *PDC1*, *ADH1*) (Supplementary 6). However, the slightly higher enzyme activities found in the YSBN2 strain (compared to CEN.PK113-7D) are not associated with higher expression of the respective genes.

Although there is no direct correlation between fluxes, enzyme activities and intracellular metabolite levels, it is striking that under batch conditions the levels of practically all amino acids were noticeably higher in CEN.PK113-7D than in the YSBN2 strain. This led us to perform a more detailed analysis of the gene expression data. Using methods for integrative analysis [129, 135] (see Supplementary 11). we calculated enriched GO-terms for the transcripts differing significantly between the two strains at both growth conditions, as well as reporter metabolites [135], and reporter transcription factors[129]. These methods allow for identification of transcriptional hot-spots in metabolic networks, i.e. metabolites around which there are large transcriptional changes, and transcription factors that drive key transcriptional responses. The results of this analysis are shown in Table 5.1. For the batch cultures the analysis points to a clear effect on amino acid transport, with several amino acid transporter genes being differentially expressed between the two strains (most transporters being expressed at a higher level in YSBN2). Also there is a distinct nitrogen catabolite repression reponse in the YSBN2 strain in the batch culture with Gln3p and Gat1p being identified as reporter transcription factors, i.e. transcription factors that controls a set of genes that have significant transcriptional changes. Thus, there are clearly differences in how the two strains control amino acid biosynthesis. YSBN2 expresses many amino acid transporters even when growing on minimal medium (without amino acid supplementation) and CEN.PK113-7D is able to maintain higher intracellular amino acid pools. The latter may be important for ensuring efficient loading of the tRNAs needed for protein biosynthesis, which in turn could be the basic explanation for the higher specific growth rate of CEN.PK113-7D.

We could not identify a direct explanation for the higher glucose uptake of CEN.PK113-7D and the differences in overall fluxes through the central carbon metabolism between the two strains. As mentioned previously several studies have shown that there is complex regulation of the glycolytic flux, and that over-expression of glycolytic genes in yeast does not result in an increased flux through this pathway[156, 163]. A hypothesis for the the higher glucose uptake of CEN.PK113-7D could, however, come from the differences in protein biosynthesis in the two strains. Protein synthesis is expensive in terms of Gibbs free energy provided in the form of ATP, and the increased glycolytic flux in the CEN.PK113-7D strain comes around from a pull of ATP and other co-factors needed for biomass production. Changes in protein biosynthesis, protein catabolism and proteolysis in the two strains could also explain the lower biomass yield of CEN.PK113-7D in chemostat culture compared with the YSBN2 strain. Thus, we find GO

Table 5.1: Overview transcriptome analysis results obtained from the comparison between CEN.PK113-7D and YSBN2 strains*

Batch				Chemostat			
GO terms[1]	Reporter metabolite	Reporter TF	ORFs	GO term	Reporter metabolite	Reporter TF	ORFs
Response to pheromone		CBK1, FUS3	FUS3, FUS1, STE5, MFA1, STE2, AGA2, BAR1, PRM10, FAR1, STE18, SST2, STE4	protein catabolism		RPN4	PIM1, PRE7, VID24, UMP1, RPN6, CDC48, RPT3, RPN9, RAD23, PRB1, RPN11, PRE4, SCL1, RPT6, UFD1, PUP2, ECM29, RPN1, PRE3, UBX6, LAP4, PRE8, PRE5, YDJ1, PUP1, RPT4, RPN8, PRE2, RPN7, SUE1
				amino acid biosynthesis	Cys, Gly, Asn, Leu	GCN4	TRP4, SAM2, GLY1, HIS1, ARG5,6, AGX1, MET13, STR3, SER2, ARG3, BAT2, ALT1, PUT1, SAM1, ECM40, LEU4, ARG1, ARG8, LEU9, ASN1
Sterol biosynthesis	(R)-Mevalonate, Episterol	MOT3	ERG28, ERG26, ERG3, HMG1, ERG13, ERG2, CYB5, MVD1	heterocycle metabolism			URA7, HEM12, TRP4, ARO10, ADE8, SAM2, HEM2, SDT1, ADE5,7, PDC6, MTD1, SAM1, HMX1, ADH6, AAH1, BIO3, ADE2, COX10
				proteolysis		RPN4	PIM1, PRE7, VID24, UMP1, PRD1, RPN6, CDC48, RPT3, RPN9, RAD23, RPN11, PRE4, SCL1, RPT6, UFD1, PHB2, PUP2, RPN1, PRE3, UBX6, PRE8, YTA12, PRE5, YDJ1, PUP1, RPT4, RPN8, PRE2, RPN7
				sterol biosynthesis	5-Diphosphomevalonate		ERG26, ERG25, ERG1, NCP1, ERG7, ERG9, ERG27, ERG6, HMG1, ERG13, ERG5, ERG2, ERG12, ERG8, CYB5, NSG2, ERG24, MVD1, **ERG20**
Amino acid transport	Leu, Phe, Ile, Val, Tyr, ALAxt, Asn, GLNxt, Cys, METxt, Trp, Gly, THRxt, PROxt, HISxt *	GLN3, GAT1	BAP3, CAN1, AGP2, **BAP2, PUT4, SAM3, DIP5, AGP1, GAP1**	allantoin metabolism	Allantoin		**DUR1,2, DAL1, DAL4, DAL2, DAL7, DAL3**
Trehalose metabolism	Trehalose 6-P, Trehalose	TPS2	**NTH1, TPS2, UGP1, TSL1, PGM2**	cytokinesis		CBK1	**BOI2, DSE1, SCW11, MYO1, DSE2, CTS1, DSE4**

*Representative GO terms, reporter metabolites, reporter TF and ORFs that were significantly changed between the two strains in the two different growth conditions (ORFs in bold indicate genes being expressed to higher levels in the YSBN2 strain relative to the CEN.PK strain; ORFs underlined indicate genes being expressed to lower levels). Complete lists are given in Supplementary 12. The representative GO terms were identified using hypergeometric tests on the genes that had significant differences in expression for the two strains. The reporter metabolites, calculated using the algorithm of Patil and Nielsen[135], indicate locations in the metabolism around which there are large transcriptional differences between the two strains. The reporter TFs, calculated using the algorithm of Oliveira *et al.*[129], indicate transcription factors for which there are significant changes in expression of the genes they are controlling. Finally the significant ORFs indicate representative examples of ORFs with significantly changed expression between the two strains. Addition of xt marks that only the extracellular amino acid was identified as a reporter metabolite, whereas for those marked without xt both the intra- and extracellular amino acid was found to be a reporter metabolite.

terms on protein catabolism and proteolysis and a large number of genes associated with these terms were significantly higher expressed in the CEN.PK113-7D strain compared with the YSBN2 strain. This is clearly linked to the transcription factor Rpn4, that is regulating the 26S proteosome. This indication of increased protein turn-over in CEN.PK113-7D is also consistent with our finding of a Gcn4 response in that strain. Gcn4 is a transcription factor that positively regulates the transcription of a large number of genes which encode for amino acid biosynthetic enzymes. In fact, a large number of these genes are also indentified as a being signifantly higher expressed in CEN.PK113-7D compared with YSBN2. Taken together, these observations strongly point to higher amino acid and protein biosynthesis, as well as increased protein degradation in CEN.PK113-7D (see Fig. 3). Since amino acid synthesis and polymerization represent the most energetically costly processes in biomass formation, a higher rate of protein turn-over in the CEN.PK113-7D strain could certainly explain the lower biomass yield. This provides a molecular explanation for the general statement made above; the CEN.PK113-7D strain has a more efficient machinery for rapid protein biosynthesis and it can thereby grow faster, but the effect of this is a less efficient utilization of the carbon and energy source at starved conditions. This is consistent with thermodynamic analysis of microbial growth, where it is found that there is trade off between the Gibbs free energy dissipation as a driving force for growth and the biomass yield[191]. The changes oberved in the CEN.PK113-7D strain are summarized in Figure 5.3, where it is illustrated that increased activity of protein catabolism and proteolysis results in the formation of a futile cycle where there is a net consumption of ATP, and this imposes a requirement for increased catabolism and respiration that ensures supply of ATP.

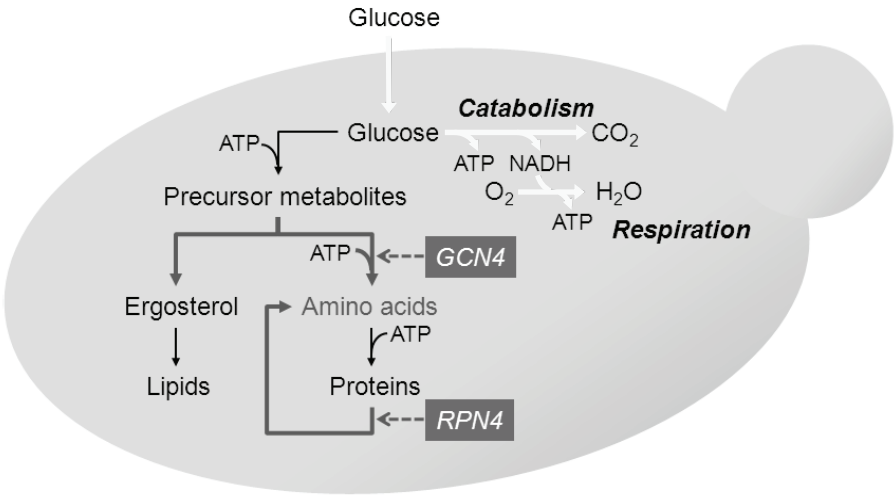


Figure 5.3: Summary of changes in metabolism in the YSBN2 strain compared with the CEN.PK113-7D strain in chemostat culture. It was observed that there is down-regulation of amino acid biosynthesis, protein catabolism and proteolysis in the YSBN2 strain compared with the CEN.PK113-7D strain (pathways that are transcriptionally down-regulated are indicated as dark bold lines). This is likely to result in less futile cycling of amino acids in the YSBN2 strain, with a reduced net consumption of ATP associated with biomass production. This results in a higher biomass yield as there is a more efficient utilization of the ATP generated in catabolism (decreased flux indicated as light bold lines) and hence relatively more glucose can be shunted towards biomass production.

Finally it can be mentioned that we observed that, at both growth conditions, there was a higher expression of genes involved in sterol biosynthesis in CEN.PK113-7D, and this is consistent with findings that the level of ergosterol is significantly higher in CEN.PK113-7D than in the S288c strain, to which the YSBN2 strain is closely related (data not shown).

In conclusion we demonstrate that through integrative analysis of a comprehensive dataset it is possible to obtain molecular explanations for observed phenotypes that would not be possible from a single omics dataset. Furthermore, our interlaboratory comparison of different methods and the detailed protocols provided, allows implementing the appropriate analytical platform in connection with systems biological studies of yeast in different laboratories. Finally, we are confident that our interlaboratory comparison of different experimental methods for omics analysis provides very useful reference data sets for two yeast strains, and these reference data sets will allow further advancement of yeast systems biology, in particularly in further use of the newly-constructed YSBN strains that represent a valuable resource for the yeast systems biology community as they are prototrophic (thus suited for physiological studies) and yet closely related with the widely used BY-strain series.

ACKNOWLEDGEMENTS

This work has been conducted in the frame of the EU-funded project "Yeast Systems Biology Network" (YSBN, contract number LSHG-CT-2005-018942) which partly funded this research. We also acknowledge the Chalmers Foundation and the Knut and Alice Wallenberg Foundation for financial support.

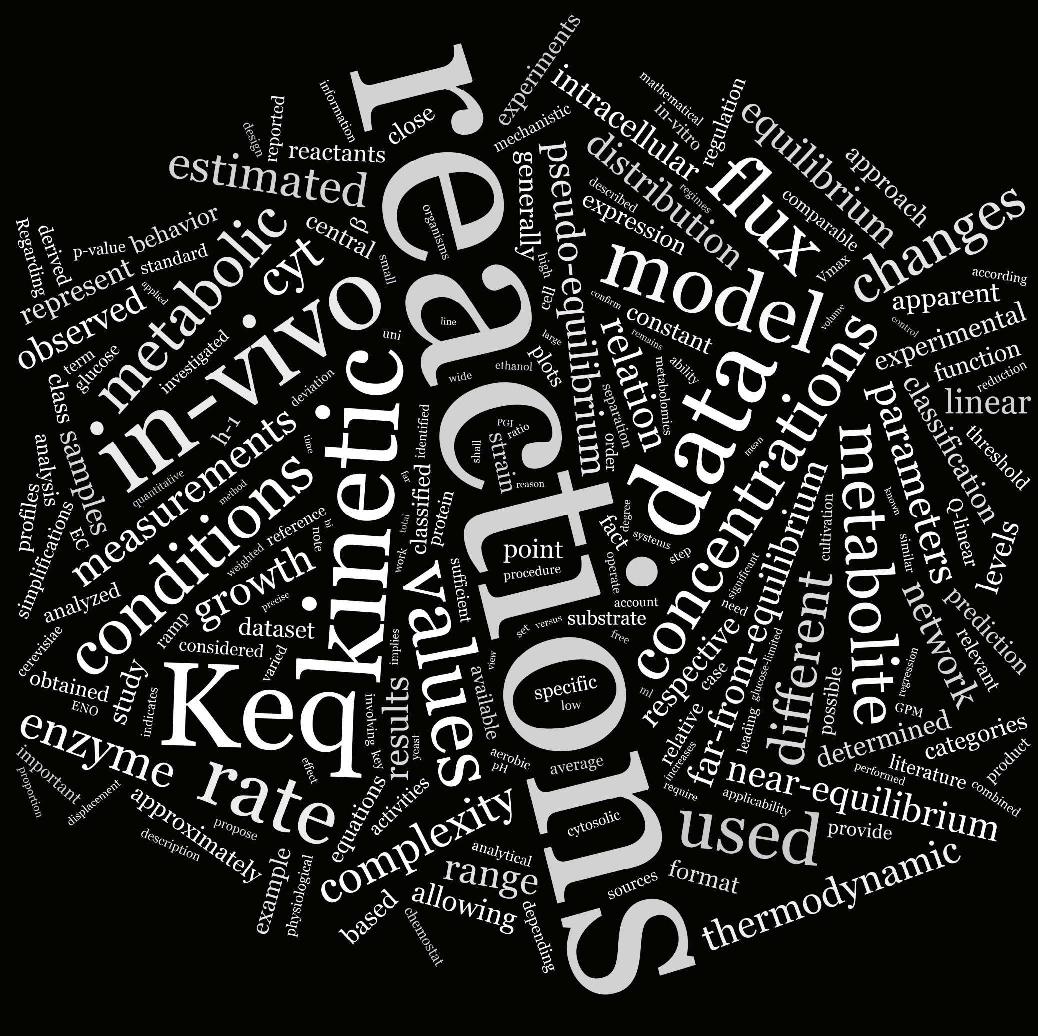
SUPPLEMENTARY MATERIAL

The supplementary material is too extensive to be integrally included in this thesis. It can be accessed online at http://www.sysbio.se/supp/.

Chapter 6

Finding simplicity in the midst of complexity

A data-driven thermodynamics-based framework for classification and quantification of *in vivo* reaction kinetics



ABSTRACT*

The construction of *in vivo* kinetic models of metabolic networks is a core ambition of Systems Biology, but has been hindered by the gap between overwhelming model complexity and limited availability and information content of *in vivo* data. To facilitate model formulation, we introduce a procedure that uses experimentally-accessible *in vivo* data to categorize reactions as pseudo-, near- or far-from-equilibrium, allowing the complexity of mathematical description to be precisely tailored to the complexity of kinetic behavior displayed *in vivo*.

We demonstrate the approach in *S. cerevisiae*, using chemostats to generate a comprehensive metabolomics dataset comprising most of central metabolism, under 32 conditions spanning flux ranges up to 60-fold. We straightforwardly derived fully *in vivo*-based kinetic descriptions, which can be directly incorporated into mathematical models, for 3/4 of the reactions analyzed. For near-equilibrium reactions this involved a convenient linear kinetic format, which we dubbed "Q-linear kinetics". We also performed, for the first time, systematic estimation of apparent *in vivo* Keq values. Remarkably, comparison with *E. coli* data suggests they constitute a suitable reference for *in vivo* thermodynamic analyses across species.

The approach is readily scalable and applicable to other biological systems.

* Manuscript in preparation: Finding simplicity in the midst of complexity: a data-driven thermodynamics-based framework for classification and quantification of *in vivo* enzyme kinetics

INTRODUCTION

An old ambition in Biochemistry and a central goal in Systems Biology is the construction of kinetic (dynamic) models of complete metabolic reaction networks, quantitatively describing the relations between fluxes, metabolite concentrations and protein levels. By allowing us to simulate complex network behavior beyond our intuitive abilities, kinetic models are valuable tools for developing and testing hypotheses regarding the regulatory circuitries controlling cellular metabolism. Furthermore, once integrated with models of gene expression and regulation of enzyme activity (from transcriptional regulation to protein interaction networks), such models may provide a powerful basis for the rational re-design of metabolic properties via genetic and protein engineering. However, the promise of kinetic modeling remains largely unfulfilled because of a basic problem in the analysis of enzyme kinetics *in vivo*: the overwhelming model complexity is unmatched by the limited availability and information content of *in vivo* data[68, 126]. This gap between model complexity and data availability generally precludes comprehensive *in vivo*-based parameterization, not to mention model discrimination. To close this gap, efforts must be aimed not only at expanding our ability to generate data, but also at developing and improving approaches to manage the complexity of kinetic models. Here we propose a data-driven approach to classify reactions according to their thermodynamic state and kinetic behavior *in vivo*. We envision such an approach as a framework to guide the initial stages of model formulation, by allowing the degree of complexity of the mathematical description of each reaction to be precisely tailored to the complexity of the kinetic behavior observed *in vivo*.

The kinetics of enzyme-catalyzed reactions are generally expressed mathematically by means of rate equations. For the most common reaction mechanisms (e.g. ping pong, ordered bi bi) the non-linear rate equations derived from mechanistic considerations contain 9-10 kinetic parameters, while for reactions involving more complex interactions (e.g. allosteric activation/inhibition) they can be even more elaborate. In principle, large-scale kinetic models could be assembled from known biochemical properties of the individual enzymes. In practice, kinetic information is incomplete or non-existent for the majority of enzymes of the majority of biological sources (e.g. strains, species, tissues), and the ability to predict quantitatively *in vivo* fluxes from *in vitro*-derived kinetic parameters is questionable [119, 155, 160, 169, 173]. Generally, the values for most (if not all) parameters are unknown and need to be estimated, for example by global fitting procedures, using *in vivo* datasets comprising fluxes, metabolite concentrations and protein levels. Here, the problem of parameter non-identifiability emerges [68]. In contrast with the overwhelming complexity of kinetic models, the information content that can be generated via *in vivo* experimentation is fundamentally restricted by the use of intact living cells. Protein and metabolite levels cannot be varied freely and independently, as in a test tube, which greatly constrains the observability of kinetic properties. Additionally, the characteristics of metabolic systems can also give rise to issues related to low parameter sensitivities [69, 194], colinearity of metabolite profiles or time-scale separation [125, 126]. Finally, simply generating sufficient data can be difficult because obtaining meaningful absolute measurements of all relevant fluxes, metabolites and proteins is still a rather demanding task.

Consequently, the fundamental constraints imposed by the experimentation with intact cells have come to be regarded as a central problem for *in vivo* kinetic analysis. Here, we propose to turn them into an advantage, by exploiting them for the purpose of eliminating excessive model complexity. The key realization which we aimed to explore was that although enzymes can in principle operate under any set of metabolite concentrations, their kinetic behavior *in vivo* may be simpler if the concentration space is highly constrained. Indeed, the fact that biochemical reactions take place within a highly connected network means that metabolite levels vary within certain physiological ranges (e.g. due to thermodynamic constraints) and that changes in their concentrations are not independent, but often correlated. If due to such constraints the concentration space that enzymes experience *in vivo* is much narrower than what they can be exposed to *in vitro*, it may not be possible to parameterize fully detailed rate equations (meant to predict behavior over wide ranges of concentrations) based on *in vivo* data. A second aspect is that the complexity of *in vivo* kinetic behavior differs from reaction to reaction. Ideally, the degree of complexity of each rate equation should be just sufficient to describe the possible range of conditions achievable *in vivo*, and not more. So the key challenge is to identify the degree of complexity required to capture the kinetic behaviour of each individual enzyme under in-vivo conditions, so that the complexity of the rate equation can be tailored accordingly. That is what the classification approach we propose here aims to accomplish.

Classification concept and thermodynamic background

We propose a reaction classification scheme as depicted in Figure 6.1, which we shall describe here in detail.

The starting point is an experimental design which should allow the state of the network to be probed over a sufficiently wide range of physiologically relevant flux regimes. A comprehensive dataset is then generated comprising, for each condition, the fluxes, the metabolite concentrations and the protein levels. Two measures derived from the experimental data are used to examine the thermodynamic state and kinetic behavior of each reaction: the order of magnitude of the measured reaction quotient (Q) in relation to (estimates of) the equilibrium constant (K_{eq}), often expressed as the displacement from equilibrium (Q/Keq); and the shape of the Q versus flux (v) profile. Based on Q/Keq and the Q(v) plot, each reaction can be classified into one of three categories: pseudo-, near- and far-from-equilibrium.

The significance of the selection criteria used is best understood by considering that, using standard notation [39], the rate of an enzyme-catalyzed reaction can be written as the product of three factors:

$$v = V_{max} \cdot \beta(x,p) \cdot \left(1 - \frac{Q}{K_{eq}}\right) \qquad \text{(equation 6.1)}$$

The first term, Vmax, is the maximal velocity in the forward direction, and represents catalytic capacity. The second term, β, is a non-linear function of affinity parameters (p) and metabolite concentrations (x) and represents the binding of reactants and effectors. Its mathematical form

depends on the reaction mechanism, so it may differ from enzyme to enzyme. The third term, 1-Q/Keq, represents the thermodynamic driving force. The link between Q and thermodynamics becomes clear if we recall that the displacement from equilibrium, Q/Keq, is directly related to the gibbs free energy of reaction, $\Delta_r G$:

$$\Delta_r G = \Delta_r G^0 + RT.\ln(Q) = RT.\ln\left(\frac{Q}{K_{eq}}\right) \qquad \text{(equation 6.2)}$$

The order of magnitude of Q/Keq provides the first criterion of classification. If the measured Q is much lower than Keq over the relevant range of physiological conditions (right-hand graph in Figure 6.1), then (1-Q/Keq)≈1 and the reaction rate is practically insensitive to changes in Q. In such cases, the thermodynamic driving force is always maximal and plays no role in the modulation of flux (other than setting its direction). Thus, the reaction rate is regulated solely by changes in enzyme capacity and binding of reactants and effectors. Such reactions shall be classified as "far-from-equilibrium" (also sometimes referred to as "irreversible").

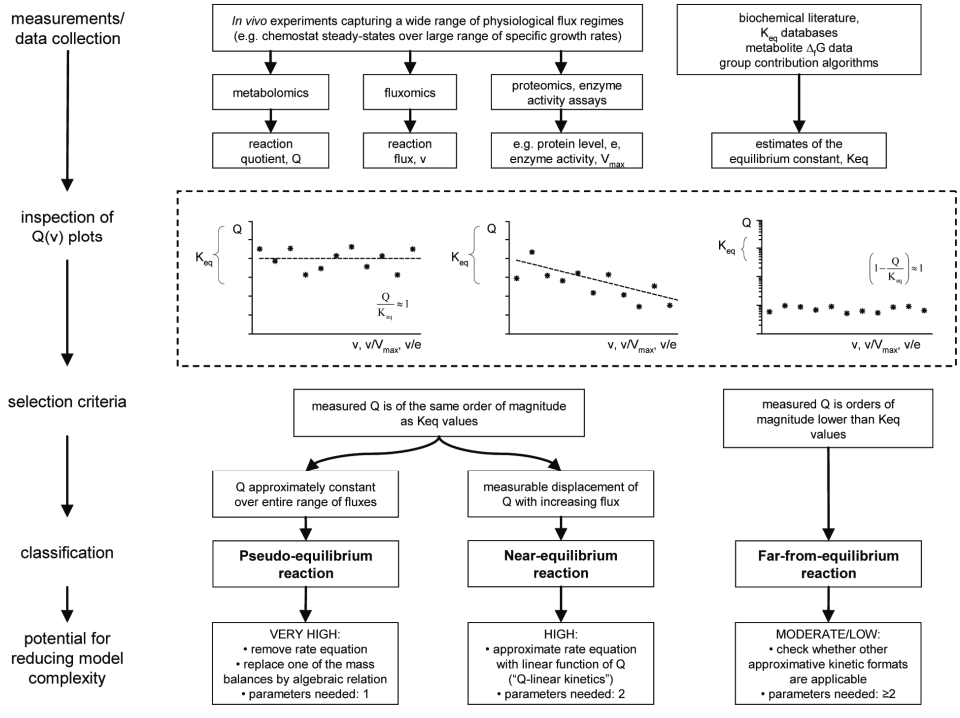


Figure 6.1: Scheme of the reaction classification procedure

To implement the identification of far-from-equilibrium reactions, we used the average Q observed at the lowest flux regimes for each reaction and compared it to the respective Keq. In this regard, it is important to note that obtaining Keq data is not yet straightforward. For most reactions data is scarce and values from different sources sometimes differ by orders of magnitude. In fact, the uncertainty in Keq data is a known problem and remains one of the main obstacles to accurate thermodynamic analysis of metabolic networks [55, 76, 99, 118]. In this study we have chosen to use the database of experimentally determined Keq compiled by Goldberg et al [66] as "golden standard", resorting to other sources only if no experimental Keq value was found. Also the choice of threshold of Q/Keq, below which a reaction will be identified as far-from-equilibrium, requires deliberation. It should take into consideration the uncertainty in Keq but also the possibility that the apparent Keq under intracellular conditions differs from the *in vitro*-determined values. For the purpose of this study we assumed that one order of magnitude should be sufficient to take these effects into account and, thus, a threshold of Q/Keq<0.1 was used as criterion for classification as "far-from-equilibrium".

The reactions which are not classified as far-from-equilibrium can be further segmented by inspection of the experimentally observed Q/v relation. To highlight the significance of this relation, the generic rate equation above can be re-arranged:

$$Q = K_{eq} \cdot \left(1 - \frac{v}{\beta(x,p).V_{max}} \right) \qquad \text{(equation 6.3)}$$

By definition, at v=0, Q=Keq, and the reaction is said to be at equilibrium. But as the flux increases, Q will depart from Keq. The exact shape of the Q/v relation as the flux increases will depend on the values of Vmax and the β function at each point. The extent of displacement of Q, over the relevant range of physiological conditions, provides the second criterion of classification. If the displacement from equilibrium is so small that it is not detectable, or not deemed statistically significant (left-hand graph in Figure 6.1), the reaction shall be classified as "pseudo-equilibrium". This would be the case, for example, if $V_{max}>>v$. If the displacement is measurable and considered statistically significant (centre graph in Figure 6.1), the reaction shall be classified as "near-equilibrium".

To implement the identification of the pseudo-equilibrium reactions, we have used the slope of the weighted linear regression of Q versus v and evaluated the p-value (of a two-tailed T-test distribution) for whether it equals zero. A high p-value corresponds to a Q(v) profile that is approximately constant. The optimal choice of the p-value threshold may vary depending on the system being studied, the quality of the measurement data, and the tradeoff between precision and simplification that is considered acceptable. For the dataset used in this study, a somewhat loose criterion of p>0.005 was used.

From a kinetic modeling perspective, far-from-equilibrium reactions present the least potential for model simplifications. These reactions represent the so-called "irreversible" steps and are thought to constitute key sites of flux regulation, for example by means of allosteric interactions. Thus, to ensure correct prediction of flux distribution it is crucial to correctly capture the role of all key effectors, which may often require the use of detailed mechanistic rate equations. It seems likely that in some cases simplifications to the rate equation will be

acceptable, for example by employing approximative kinetic formats [74], but their adequacy may need to be examined for each system on a case-by-case basis [27].

In contrast, the reactions classified as pseudo-equilibrium reactions allow major reductions in complexity, which makes their accurate identification highly advantageous from a model reduction point of view. For each reaction, the rate equation can be removed and one of the metabolite mass balances can be replaced by an algebraic relation (e.g. [product]= K_{eq}.[substrate]).

The near-equilibrium reactions cannot afford such drastic simplification in their kinetic description, thus representing an intermediate level of complexity. However, in the course of this work we found that they also hold considerable potential for reduction of complexity, which has not been sufficiently exploited before. As we shall demonstrate, their kinetics are mainly regulated by the thermodynamic driving force and the reaction rate is best described as a linear function of Q: a kinetic format which we shall refer to as "Q-linear kinetics" (see Results and Discussion). This allows replacing enzyme-specific non-linear rate equations, typically containing 4 to 10 parameters, by unified linear format ("Q-linear kinetics") containing only 2 parameters which can be accurately determined from *in vivo* data.

Experimental design

The proposed classification procedure entails the investigation of the kinetic behavior of individual reactions in great detail, which requires a comprehensive set of data: the concentrations of all reactants, the fluxes and the enzyme capacities. And, as we have emphasized, the success of the approach relies partly on capturing a sufficiently wide and physiologically relevant range of flux regimes, which implies sampling at many growth/cultivation conditions. Metabolic Flux Analysis is a relatively well established field and allows the estimation of intracellular flux distributions. Metabolomics has reached a stage, in terms of accuracy and network coverage, which allows the absolute quantification of most intermediates of central metabolism and certain anabolic pathways. However, reproducible absolute determination of protein levels is not yet accessible to most laboratories and assaying enzyme activities across entire pathways would be extremely time-consuming, and may not always be possible. So the ideal situation to test the procedure we propose would be the somewhat counter-intuitive scenario where fluxes would change by severalfold but enzyme levels would not.

Fortunately, such a situation is known: aerobic glucose-limited chemostat cultivations of *Sacharomyces cerevisiae* at different specific growth rates. It has been observed that under such conditions large changes in growth rate, leading to huge changes in flux through the metabolic network (e.g. >20-fold), are associated with comparatively small changes (e.g. <2-fold) in gene expression [144] and enzyme capacities [47, 178]. Additional evidence that responses to environmental perturbations typically involves relatively small changes in enzyme activities [42, 43, 139, 173] also point to a general pattern of dominant metabolite-level regulation of flux in the central metabolism of yeast. Thus, by allowing fluxes to be easily controlled in a well-defined manner, over a very wide range, with minimal changes in enzyme levels, aerobic glucose-

limited chemostat cultures of *S. cerevisiae* provide an ideal system for studying the relations between fluxes and metabolites.

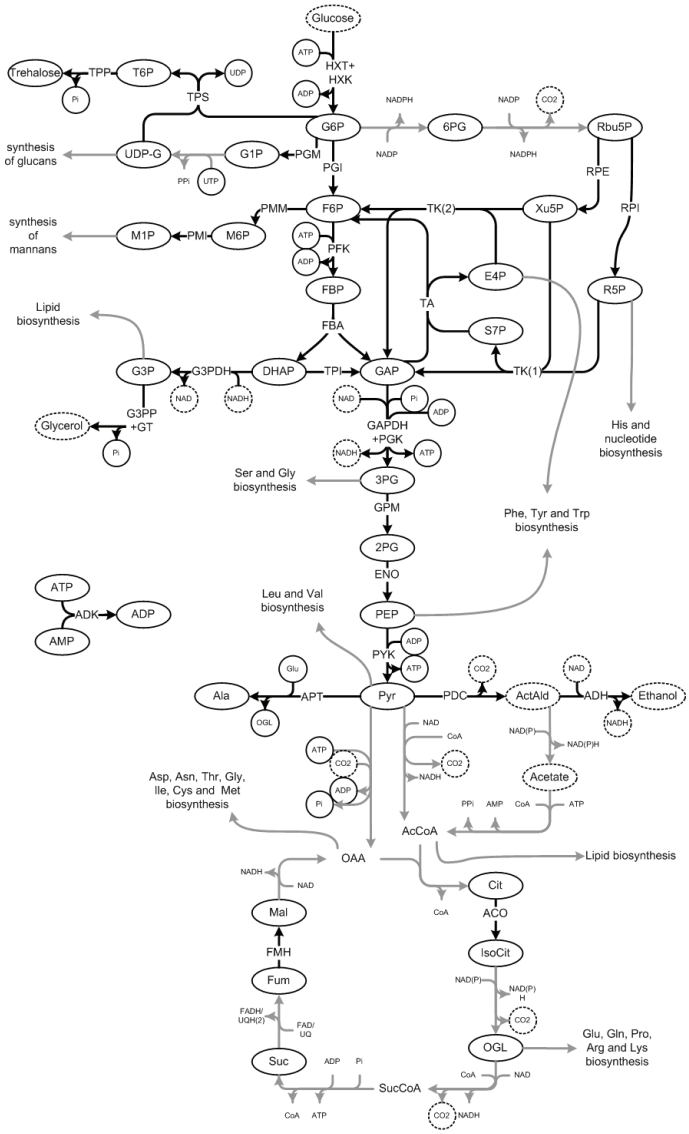


Figure 6.2: Diagram of the reactions analyzed in the central metabolism of yeast. Circles represent measured metabolites; circles in dashed line represent metabolites which are measured extracellularly (e.g. glucose, by-products) or indirectly estimated (cytosolic free NAD/NADH ratio); metabolites without circle are unmeasured (e.g. mitochondrial coenzymes). Arrows in black represent reactions which can be analyzed in detail (all reactants are measured); arrows in gray represent those which cannot; arrowheads indicate flux directions under the experimental conditions investigated (aerobic, glucose-limited). See text for abbreviations.

The experimental design we have chosen to implement was inspired by the work of van Hoek et al. [178]. With a combination of experiments involving dilution rate steps and ramps, a total of 32 well-defined growth conditions were investigated, covering a range of growth rates from 0.02 to 0.38 h^{-1} (μ_{max}=0.4 h^{-1}). Intracellular flux distributions were determined using a validated

stoichiometric model [42] and a comprehensive quantitative metabolomics dataset was obtained using well-validated sample preparation techniques [31, 32] and a set of four complementary targeted IDMS-based analytical platforms [32, 37, 158, 172]. The resulting dataset allows the detailed kinetic/thermodynamic analysis of 27 reactions, representing most of the central metabolism in yeast (Figure 6.2).

MATERIALS AND METHODS

Solvents and chemicals – HPLC-grade solvents were from Mallinckrodt Baker. Analytical grade standards were from Sigma-Aldrich.

Strains and cultivation conditions – Two closely related *S. cerevisiae* strains were used. The commonly used CEN.PK 113-7D strain [176], also employed by van Hoek et al.[178], was used for the step experiments (see below). The "mtlD" strain, which expresses a reporter system for the cytosolic free NAD/NADH ratio [33], was used in the ramp experiments (see below). Overall physiology (exchange rates and yields) and the metabolite measurements confirm that under the same conditions the behavior of the metabolic network is comparable in the two strains, confirming that the datasets could be combined. Low-salt defined medium according to Canelas et al.[32] was used, with 7.5 g/L glucose as carbon source and 0.05 g/L silicone antifoaming agent (BDH, UK). The cells were grown under aerobic carbon-limited chemostat conditions, in a 7 L bioreactor (Applikon, The Netherlands) equipped with a DCU3 controller and continuous data acquisition via MFCS (B. Braun Biotech, Germany). The cultivations were carried out with a working volume of 4 L, at 30°C, pH 5.0, 0.3 bar overpressure, an aeration rate of 0.5 vvm and a stirrer speed of 500 rpm. Standard fermentation and analytical procedures were performed as described earlier [32], unless stated otherwise.

Chemostat cultures – The design of the experiments were inspired by the work of van Hoek et al.[178]. Two "step" experiments were performed with CEN.PK 113-7D, each investigating four specific growth rates, in tandem (0.02-0.1-0.3-0.35 and 0.05-0.2-0.32-0.38 h^{-1}). Each cultivation was initiated at the lowest growth rate. After at least 2 residence times of steady-state growth with constant dissolved oxygen and exhaust-gas readings, two samplings were performed, separated by 2-6 h, for intracellular and extracellular metabolite determination. After the second sampling, the dilution rate was increased to the next set-point, and so on until four conditions had been sampled. The effective dilution rate, continuously monitored via the weight of the effluent vessel, was routinely checked and corrected as necessary. The total length of the cultivations was purposefully kept to a minimum (45-50 generations) to prevent potential changes in enzyme activities caused by long-term adaptation to chemostat conditions [86, 113].

Additionally, two "ramp" experiments were performed with the mtlD strain [33], investigating a total of 16 specific growth rates: ramp down from 0.1 to 0.03 h^{-1} and ramp up from 0.1 to 0.29 h^{-1}. Each experiment was initiated from a chemostat cultivation at D=0.1 h^{-1}. After at least 2 residence times under steady-state conditions, samples were taken for intracellular and extracellular metabolite determination and a linear feeding profile (downwards or upwards) was

started. Samples were taken at regular intervals for a period of 4 (ramp down) to 12 h (ramp up). These ramp experiments were designed to accomplish a wide range in growth rates, as the step experiments, but requiring a shorter total cultivation time. This was meant to minimize the chance of loss of the reporter gene (carried by a plasmid) due to selective pressure. High, sustained activity of the reporter enzyme was confirmed by *in vitro* enzyme activity measurements as well as *in vivo* perturbation response tests (change in Mtl1P within 10 s of a glucose pulse [33]). According to the data generated, the four experiments (two steps with the reference CEN.PK 113-7D strain and two ramps with the mtlD strain) resulted in overlapping metabolite versus flux patterns, confirming that all the results could be combined and treated as a single dataset.

Stoichiometric model and flux calculations – Intracellular flux distributions were calculated from the measured volumetric rates of biomass growth and production/consumption of glucose, ethanol (after correction for evaporation), O_2, CO_2, acetate and glycerol, using a stoichiometric model based on [42], with modifications to the carbohydrate synthesis pathways and biomass composition. Instead of a single lumped reaction for carbohydrate synthesis from G6P, ten reactions leading from the precursors G6P and F6P to trehalose, glycogen and structural glucans and mannans were introduced:

Phosphoglucomutase:
1 G6P:cyt <=> 1 G1P:cyt
UTP G1P uridylyltransferase:
1 G1P:cyt + 1 UTP:cyt => 1 PPi:cyt + 1 UDPGlu:cyt
Trehalose-6-phosphate phosphatase:
1 H2O:cyt + 1 T6P:cyt => 1 Pi:cyt + 1 Treh:cyt
Trehalose-6-phosphate synthase:
1 G6P:cyt + 1 UDPGlu:cyt => 1 H:cyt + 1 T6P:cyt + 1 UDP:cyt
Glycogen synthase:
1 UDPGlu:cyt => 6 Glyc:cyt + 1 H:cyt + 1 UDP:cyt
Glucan synthase:
1 UDPGlu:cyt => 6 glucan:cyt + 1 H:cyt + 1 UDP:cyt
M6P isomerase:
1 F6P:cyt <=> 1 M6P:cyt
Phosphomannomutase:
1 M1P:cyt <=> 1 M6P:cyt
M1P guanylyltransferase:
1 GTP:cyt + 1 M1P:cyt => 1 GDPMan:cyt + 1 PPi:cyt
Mannan synthase:
1 GDPMan:cyt => 1 GDP:cyt + 1 H:cyt + 6 mannan:cyt

Additionally, since the macromolecular biomass composition varies with growth rate, eight different biomass compositions were defined: 0.02, 0.05, 0.1, 0.2, 0.3, 0.32, 0.35 and 0.38 h^{-1} (for the ramp experiments the closest value was used). The biomass composition at different growth rates was statistically reconstructed, based on the method reported earlier [101], from multiple sources of data on the growth rate dependence of different biomass components for *S. cerevisiae* under aerobic conditions [50, 98, 101, 184]. Flux calculations were implemented in MNA

PR-3.0 (SPAD it, Nijmegen, The Netherlands). For simplicity, we have refrained from estimating the uncertainties in the calculated flux values.

Intracellular metabolites – Broth samples (1.20 ± 0.03 g), in duplicate or triplicate, were taken with a leakage-free cold methanol quenching method [31], via a custom-made low dead-volume rapid sampling setup [100]. Boiling ethanol extraction, vacuum evaporation and further conditioning were performed as described earlier [32]. The concentrations of G6P, F6P, G1P, M6P, M1P, Mtl1P, FBP, F2,6BP, T6P, UDP-G, 6PG, S7P PEP, G3P, PYR, OGL, SUC, FUM and MAL were determined by anion-exchange LC-MS/MS [172]. The concentrations of G6P, F6P, RBU5P, R5P, X5P, S7P, E4P, GAP, DHAP, 3PG, 2PG, CIT, ISOCIT, OGL, FUM, MAL and Trehalose were determined, as methoxime-trimethylsilyl derivatives, by GC-MS [37]. The concentrations of cAMP, AMP, ADP, ATP, UMP, UDP and UTP were determined by ion-pair reverse-phase LC-MS/MS [158]. The concentrations of 20 amino acids were determined, after SPE and derivatization with propyl chloroformate, by GC-MS [32]. All analyses were performed at least in duplicate. Quantification was based on IDMS [117, 209]. For those metabolites measurable by more than one platform, all measurements were combined.

Intracellular phosphate – For compounds present at high concentrations in the medium (≈27 mM in the case of phosphate) the method above, which involves cell separation by centrifugation, would lead to gross overestimation of the intracellular levels due to partial carry-over of the (orders of magnitude larger) extracellular fraction [31, 210]. Thus, a modified version involving rapid filtration was designed for improved washing efficiency. Using the same custom-made low dead-volume rapid sampling setup (pre-calibrated and with the vacuum system off), broth samples (2.52 ± 0.03 g) were drawn into tubes containing 30 ml pure methanol at -40ºC. The mixture was immediately poured onto membrane disk filters (PES, 0.45 µm, Pall, USA) in a vacuum filtration setup. Immediately after all the liquid passed the filter, an additional 30 ml of 90% (v/v) methanol solution at -40ºC were added for washing (from the sample tube, to wash any left-over sample there). Immediately after the washing was finished, the filter was rapidly collected using a pair of metal tweezers and submerged in 30 ml 75% (v/v) ethanol at 70ºC, the tube was closed, vigorously shaken and placed in a water bath at 95ºC. After 3 min the tube was placed back at -40ºC. The period from sampling to beginning of extraction was less than 30 s. The total turn-around time of the sampling setup (including washing of the filter holder) was 1-1.5 min. The high concentrations of methanol and short exposure time were meant to prevent leakage of the intracellular components [31]. On the other hand, tests with salt solutions confirmed there was no precipitation of phosphate and the fact that intracellular phosphate levels were comparable with literature data suggests that the washing was efficient enough. To the ethanol extracts, ^{18}O-labeled phosphate was added as IDMS internal standard. The samples were dried under vacuum, resuspended in 5 ml demineralized water and stored at -80ºC until analysis. The concentration of Pi was determined by anion-exchange LC-MS/MS [172], with quantification via ^{18}O-based IDMS. We also attempted to determine pyrophosphate, which could be separated and identified in standards, but was too low for proper quantification in samples. Intracellular phosphate was only determined for the step experiments. Since the levels appeared unaffected by growth rate, their average and standard deviation were used for the dataset of the ramp experiments.

Cytosolic free NAD/NADH ratio – The cytosolic free NAD/NADH ratio was determined in the ramp experiments with the mtlD strain, from the concentrations of F6P and Mtl1P and the equilibrium constant of M1PDH, as described earlier [33]. For the step experiments the cytosolic free NAD/NADH was estimated by linear regression versus the profiles of PEP concentration (R^2=0.96) and G3PDH flux (R^2=0.93).

Extracellular metabolites - Extracellular samples, in duplicate, were obtained by rapid filtration of 2-2.5 g broth via the cold steel beads method [116], using syringes loaded with 25.2 g steel beads pre-cooled to -20ºC and disposable glycerol-free cartridge filters (Millex, 0.45 µm, PVDF, Millipore). Residual glucose, glycerol and acetate were determined using enzymatic assay kits (R-Biopharm, Germany), scaled down to 200 µl for use in 96-well plates, via NADH absorbance at 340 nm, on a GENios microtiter plate reader (Tecan, Switzerland). Ethanol and glucose were determined by HPLC (Bio-Rad Aminex HPX-87H column, at 60°C, 0.6 ml/min 10 mM phosphoric acid as eluent, coupled to a Waters 2410 RI detector and a Waters 2487 UV detector at 214 nm). Acetaldehyde and ethanol were determined by GC (HP Innowax 19095N-121 column, at 70ºC, with a 0.2 µl injection volume and 180ºC injector temperature, helium at 60kPa as carrier gas, and coupled to a FID detector at 200ºC). For those metabolites measurable by more than one method, all measurements were combined.

Thermodynamic data – Data on Keq was primarily derived from the NIST database of Keq data compiled from biochemical literature by Goldberg and Tewari [66]. The available values for each reaction were averaged. When sufficient values were available over very different conditions and a strong influence of assay conditions was detected, only those close to physiological conditions were selected (e.g. pH=7±1, T=303±15 K, $[Mg^{2+}]$=0.1-1mM). For two reactions, PDC and TPP, Keq data was not available. The corresponding values were obtained, respectively, from the metabolite Δ_fG data assembled by Alberty [4, 5] and the dataset of Δ_rG estimated using a recently improved group-contribution algorithm by Jankowski et al [85]. These two data sources refer to pH=7, I=0.25M, T=298 K.

Specific cell volume and intracellular pH – For the same strain and identical cultivation conditions, van Eunen et al [177] found that the specific cell volume was the same under fully respiratory (μ=0.1 h^{-1}) and respiro-fermentative conditions (μ=0.35 h^{-1}). The value derived from their data was 1.5±0.1 ml/g_{DW}. A nearly constant specific cell volume is compatible with the findings of Brauer et al with a similar strain, for growth rates between 0.05 and 0.3 h^{-1} [20]. Cell size was also investigated for the same strain and growth conditions (μ=0.1 h^{-1}) by Cipollina et al [38]. The value derived from their data was 1.9±0.1 ml/g_{DW}. We have adopted a constant average value of 1.7 ml/g_{DW} for conversion of intracellular metabolite levels to concentrations in mol/l. The cytosolic and mitochondrial pH has been estimated for the same strain background under several growth conditions by Orij et al [134]. They found cytosolic and mitochondrial pH values of, respectively, 7.0-7.2 and 7.5 under glucose excess, 6.9 and 7.3 under glucose-limitation (μ=0.1 h^{-1}), 7.1 and 7.3 on galactose (respiro-fermentative carbon source), and 6.8 and 7.26 on ethanol/glycerol (non-fermentative carbon sources). There seems to be a slight dependency on metabolic regime, but the extent of change is too small to affect the equilibrium of most reactions. In the case of the dehydrogenases (GAPDH, ADH, G3PDH) the effect could nevertheless be relevant because a proton participates as reactant,

but that effect would be taken into account by the (pH-independent) reporter system for the cytosolic NAD/NADH ratio [33] (see above).

RESULTS AND DISCUSSION

Classification of the reactions in the central metabolism of yeast

As outlined above, the classification approach proposed relies on the generation of a comprehensive dataset covering a large range of flux regimes. Here, we demonstrate the application of the procedure for the study of the central metabolism of *S. cerevisiae* under aerobic glucose-limited conditions. We have used chemostat cultivations to cover a total of 32 well-defined growth conditions, over a range of specific growth rates between 0.02 and 0.38 h^{-1} (μ_{max}=0.4 h^{-1}). At each condition the balances of carbon and degree of reduction closed satisfactorily (±5%) and the overall physiology in terms of specific consumption/production rates was very comparable to previous studies, including the work of van Hoek et al. [178], on which the experimental design was inspired. Regarding the metabolite measurements, it is particularly important to ensure dense coverage of the relevant part of the network because each reaction can only be analyzed in detail if all reactants are measured. In fact, this is highly desirable for any study of *in vivo* kinetics/thermodynamics, because in the absence of complete datasets one must rely on global fittings. Using a set of four complementary targeted metabolomics platforms we obtained measurements for approximately 60 metabolites, achieving nearly complete coverage of the intermediates in glycolysis and the pentose phosphate pathway (PPP), and allowing the classification procedure to be applied to 27 reactions (Figure 6.2). Three of these were not individual reactions but pairs of consecutive reactions (HXT+HXK, GAPDH+PGK and G3PP+GLT), because their common intermediate was not measured (intracellular glucose, 1,3-BPG and intracellular glycerol, respectively). From the metabolite data, we found that the concentrations of most metabolites changed by at least 3-fold over the range of conditions investigated. Two notable exceptions were ATP and Pi, which varied by less than 40%. Among the reactants of the 27 reactions analyzed, the median change in intracellular concentration was 5-fold. For the classification, it is also important that the flux through the reactions of interest be varied over a sufficiently wide range, to ensure an adequate observation of the Q/v relation. The experiments carried out allowed the exploration of a range in specific glucose uptake rate of 42-fold. Among the 27 reactions of interest, the median change in intracellular flux was 35-fold.

With this data we have generated the Q(v) plots and applied the two selection criteria outlined above to classify each of the reactions. The first criterion regards the order of magnitude of Q/Keq, at the lowest flux regimes. The results obtained are depicted in Figure 5.3 (panel A). From this plot, the separation between categories (pseudo+near-equilibrium and far-from-equilibrium) appears unambiguous at a Q/Keq between 10^{-2} and 10^{-1}, so the threshold used seems sufficient to account for differences between the Keq *in vivo* and the corresponding *in vitro* values. This analysis leads to the classification of the 7 reactions with Q/Keq<10^{-1} as far-from-equilibrium. The respective Q(v) plots are presented in Figure 6.4.

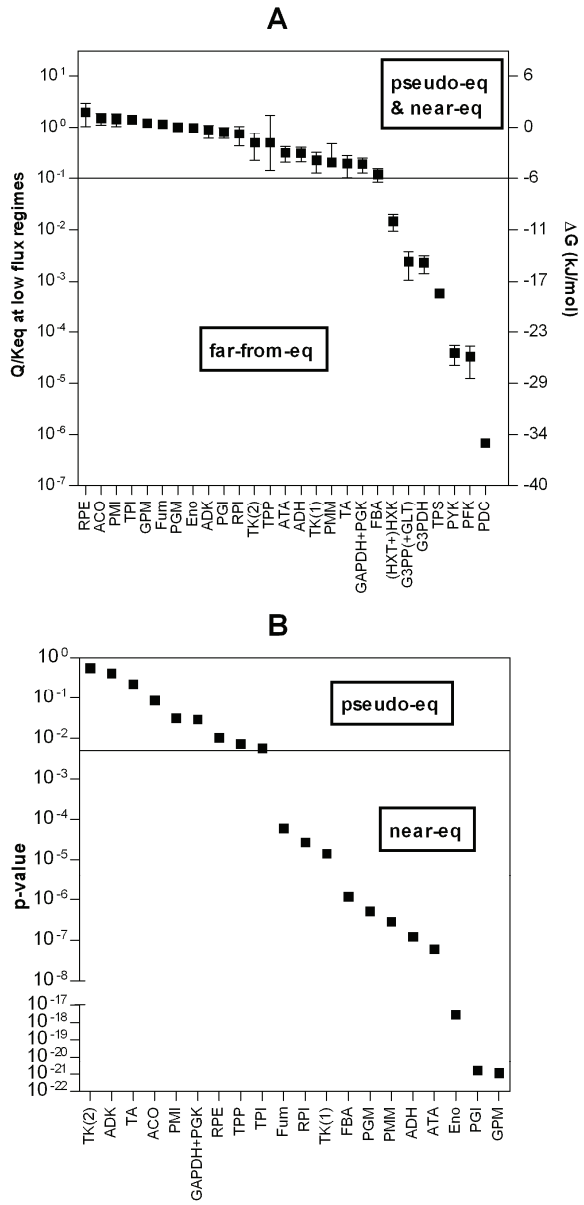


Figure 6.3: Sorting of reactions according to selection criteria. A: identification of far-from-equilibrium reactions according to Q/Keq at low flux regimes, with a threshold of 10^{-1} (horizontal line). This threshold aims to take into account both the uncertainty in Keq and the possibility that the apparent Keq under intracellular conditions differs from the *in vitro*-determined values. B: separation of pseudo- and near-equilibrium reactions according to the p-value (of a two-tailed T-test distribution) for whether the slope of the weighted linear regression of Q versus v equals zero, with a threshold 0.005 (horizontal line). A high p-value indicates that the Q(v) profile is approximately constant.

Among the far-from-equilibrium reactions we find HXT&HXK, PFK, PYK and PDC, which are commonly thought to represent some of the key sites of regulation of glycolysis and alcoholic fermentation. The classification of G3PP&GLT is also consistent with an active control of glycerol synthesis and export under aerobic conditions. Interestingly, G3PDH is the

only of the three dehydrogenases that operates far from equilibrium. Together with the fact that it is the first step towards both glycerol formation and the G3P redox shuttle, this suggests an important role of G3PDH in the control of the cytosolic redox state. Note that this might not have been detected without the recently developed reporter system for the cytosolic free NAD/NADH ratio [33], which highlights the importance of compartment-specific data in the analysis of eukaryotic systems.

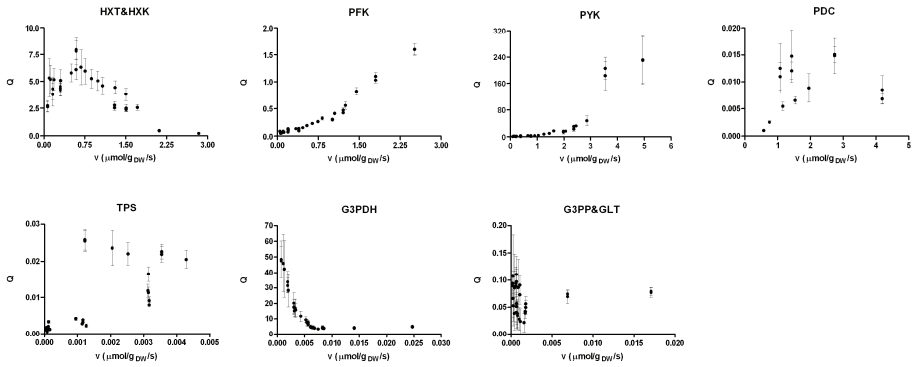


Figure 6.4: Q(v) plots for the reactions classified as far-from-equilibrium Values are averages (±standard deviation) of two samples, each analyzed at least twice. When relevant, the units of concentration are mol/l.

The rate of far-from-equilibrium reactions is practically unresponsive to Q because, if Q<<Keq, the value of the thermodynamic driving force term (1-Q/Keq) becomes insensitive to changes in Q. Therefore, we would predict that there need not be any direct relation between v and Q, and the Q/v profiles could have any shape or trend. This is indeed what we observed (Figure 6.4). For example, the most striking illustration of lack of correlation between Q and v is that increases in flux can be associated with simultaneous accumulation of product and depletion of substrate, as observed with PFK and PYK. Clearly, when changes in the respective enzyme activities are known to be small [178], such profiles imply the existence of more complex mechanisms of regulation, such as allosteric control or post-translational modification. In the case of PFK and PYK, the changes in flux can be explained by the large increases in concentrations of the respective allosteric activators F26bP and FBP. Thus, even prior to a more detailed kinetic analysis, the inspection of the Q(v) plots can already point to particular patterns of regulation

From the point of view of kinetic model complexity, an important implication of the fact that reactions in this category will exhibit more intricate kinetics is that more detailed rate expressions may be necessary to adequately describe their behavior. In particular, the interactions with effectors need to be correctly captured, as they are expected to play a major role in defining the properties and behavior of the system. This does not necessarily mean that there is no scope for further model reduction. In fact, further rate simplifications will be necessary whenever parameter identifiability issues arise due to the complexity of full

mechanistic expressions being unmatched by the availability of *in vivo* system information [126]. We envision a strategy where the initial round of reaction classification is followed by a second round of model reduction based on the use of approximative kinetic formats [74], focusing on the far-from-equilibrium reactions. However, it is important to note that which format achieves the best performance, in terms of predictive ability and parameter identifiability, will differ from reaction to reaction [27].

Having defined the far-from-equilibrium reactions, the remaining two classes can be distinguished using the second criterion, which relates to the shape of the Q(v) plot. The sorting of the reactions according to the p-value (of whether the slope of the weighted linear regression is zero) is depicted in Figure 6.3 (panel B). In this dataset, the separation between categories appears unambiguous at p-values around 10^{-4} to 10^{-3}, so a somewhat loose threshold of 0.005 seems justified. This leads to the classification of 9 reactions as pseudo-equilibrium and 11 reactions as near-equilibrium. The respective Q(v) plots are presented in Figure 6.5 and Figure 6.6.

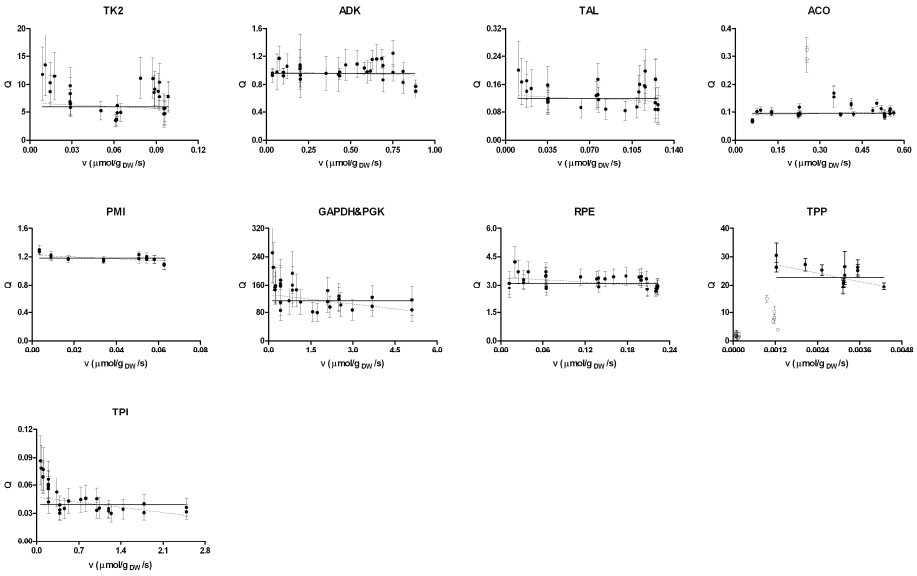


Figure 6.5: Q(v) plots for the reactions classified as pseudo-equilibrium The bold horizontal line represents the weighted average of Q. For comparison, the dotted line represents the weighted linear regression. Hollow circles (in the profiles of ACO and TPP) represent data points left out due to indications of significant changes in enzyme capacity (see Results and Discussion). Values are averages (±standard deviation) of two samples, each analyzed at least twice. When relevant, the units of concentration are mol/l.

Regarding the p-value threshold, it is worth considering that a more typical threshold of 0.01or 0.05 could also have been used, implying a more stringent criterion and resulting in less reactions being classified as pseudo-equilibrium (6 or 4, respectively). From a modeling perspective, the choice of the threshold should reflect whether the priority is to minimize model complexity or maximize predictive ability. Furthermore, the order of magnitude of the

p-values, as well as their distribution, can be expected to differ substantially depending on the number of data points and the quality of the data. Therefore, the adequacy of different thresholds should be evaluated for each dataset, for example by visual inspection of the Q(v) plots.

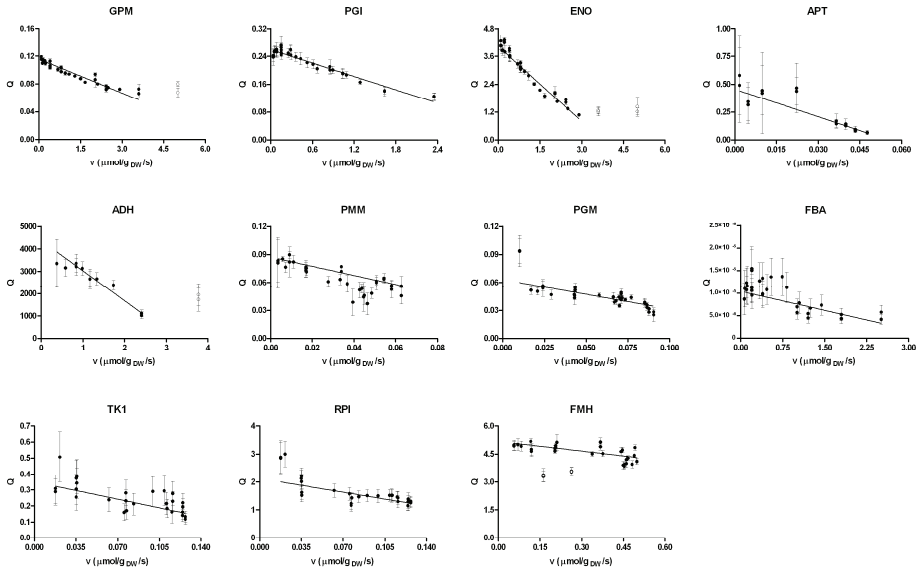


Figure 6.6: Q(v) plots for the reactions classified as near-equilibrium The line indicates the weighted linear regression. Hollow circles (in the profiles of GPM, ENO, ADH and FMH) represent data points left out due to indications of significant changes in enzyme capacity (see Results and Discussion). Values are averages (±standard deviation) of two samples, each analyzed at least twice. When relevant, the units of concentration are mol/l.

Interestingly, among the reactions classified as near-equilibrium we found several reactions which have previously been reported to behave as pseudo-equilibrium, such as PGI, GPM, ENO, PGM and FMH [96, 112, 189]. Previous studies of *in vivo* kinetics were mostly based of short-term stimulus response data (e.g. glucose pulse to a glucose-limited chemostat). Their inability to observe displacement from equilibrium can be attributed to the fact that even a glucose pulse to a low specific growth rate culture can only achieve a 3 to 6-fold change in uptake rate, which is relatively small compared to the range of fluxes accomplished in this study. Among the reactions classified as pseudo-equilibrium we found several reactions which, in the absence of measurement data, are commonly assumed to be pseudo-equilibrium in modeling studies, such as ADK, TPI, TK2 and TAL. Based on the results we can confirm that this assumption is indeed justified in the case of *S. cerevisiae* under aerobic glucose-limited conditions.

It is essential to realize that the ability to effectively distinguish pseudo- from near-equilibrium reactions stems directly from exploiting the Q/v relation. Theoretically, it could be much easier to identify pseudo-equilibrium reactions, by quantitatively comparing Q and K_{eq}. But in practice, that approach is undermined by the limited availability and the high degree of

uncertainty in the K_{eq} data, which generally exceeds 2-fold, as well as the possibility that the Keq under *in vivo* conditions differs from the *in vitro*-determined Keq. It is only via the procedure outlined here, of employing the Q(v) plots, that the need for precise Keq data can be circumvented, thus enabling the separation of the two classes.

Regarding the reactions classified as pseudo-equilibrium, it is important to keep in mind that over the range of flux regimes investigated the reaction rates have changed by 9 to 65-fold and the concentrations of most substrates and products have changed by 2 to >10-fold. Thus, the fact that Q remains constant can be considered a true system property, not a consequence of insufficient stimulation.

For modeling purposes, the identification of pseudo-equilibrium reactions is notably useful because of the powerful model simplifications that it allows: the respective rate equation can be removed from the model and one mass balance (ODE) can be replaced by an algebraic equilibrium relation. This way, all parameters are eliminated but one, Keq, which implies a radical reduction in model complexity. It is important to realize that the reason that the capacity and affinity parameters can be eliminated is that they have no observable effect in the *in vivo* kinetic behavior of the network.

Near-equilibrium reactions can be conveniently described via "Q-linear kinetics"

The most striking feature observed in the Q(v) profiles was how well the relation between Q and v could be approximated by a linear function in the case of the near-equilibrium reactions**,** over such wide ranges of fluxes. This had not been anticipated because in principle the relation between Q and v can vary depending on the changes in V_{max} and the β function.

Regarding V_{max}**,** its value was indeed expected not to change much for most enzymes over the range of conditions investigated [178], but changes some enzymes cannot be excluded. For example, deviations observed in the Q(v) profiles for GPM, ENO and ADH (Figure 6.6) can be attributed to increased expression under fermentative conditions [178] (an increased activity corresponds to a decreased slope). Also the deviations in TPP (Figure 6.5) can be explained by the tight regulation of trehalose synthesis above intermediate growth rates, which is known to involve active protein degradation [50, 57, 161] (the data points cluster towards the origin because the flux also decreases; notice a similar shut-down for TPS). Only in the case of FMH and ACO, it cannot be clarified whether the deviations are caused by changes in gene expression [29], enzyme localization [145], or the compartmental distribution of their substrates. As a side note, it is interesting to observe that, although that was not the main focus of this work, the consequences of regulation events could be detected in the metabolite data, demonstrating the potential of quantitative metabolomics for revealing phenotype and its importance for integrated analysis of metabolic regulation systems.

Thus, if V_{max} can be considered relatively constant for most enzymes over most of the flux range, the nearly linear relation between Q and v implies that the value of the function β, which represents the interactions of the enzyme with reactants and effectors, must also be approximately constant. This raises the question: why should the value of β, which is generally

a complex non-linear function of affinity parameters and metabolite concentrations, remain relatively constant when the metabolite concentrations are changing by 3 to >10-fold? A mechanistic explanation was found in the properties of the β function and the fact that concentrations of reactants in near-equilibrium reactions do not, by definition, change independently. The mathematical form of the function β depends on the reaction mechanism. Examples of the rate equations for the most common (non-allosteric) mechanisms are presented in Table 6.1.

Table 6.1: Examples of rate expressions for typical reaction mechanisms. Note that PGM, GPM and ENO are bi bi reactions with regenerated cofactors (glucose-1,6-bisphosphate, 2,3-diphosphoglycerate and Mg^{2+}, respectively).

Uni uni (e.g. PGI, RPI, TPI):

$$v = V_{\max} \frac{\frac{a}{K_a}}{1 + \frac{a}{K_a} + \frac{p}{K_p}} \left(1 - \frac{\frac{p}{a}}{K_{eq}} \right)$$

Ordered uni bi (e.g. FBA):

$$v = V_{\max} \frac{\frac{a}{K_a}}{1 + \frac{a}{K_a} + \frac{K_q \cdot p}{K_{iq} K_p} + \frac{q}{K_{iq}} + \frac{ap}{K_a K_{ip}} + \frac{pq}{K_p K_{iq}}} \left(1 - \frac{\frac{pq}{a}}{K_{eq}} \right)$$

Ping-pong (e.g. APT, TK1, TK2, PGM, GPM):

$$v = V_{\max} \frac{\frac{ab}{K_{ia} K_b}}{\frac{a}{K_{ia}} + \frac{K_a b}{K_{ia} K_b} + \frac{p}{K_{ip}} + \frac{K_p \cdot q}{K_{ip} K_q} + \frac{ab}{K_{ia} K_b} + \frac{ap}{K_{ia} K_{ip}} + \frac{pq}{K_{ip} K_q} + \frac{K_a bq}{K_{ia} K_b K_{iq}}} \left(1 - \frac{\frac{pq}{ab}}{K_{eq}} \right)$$

Ordered bi bi (e.g. ADH, ENO):

$$v = V_{\max} \frac{\frac{ab}{K_{ia} K_b}}{1 + \frac{a}{K_{ia}} + \frac{K_a b}{K_{ia} K_b} + \frac{K_q \cdot p}{K_{iq} K_p} + \frac{q}{K_{iq}} + \frac{ab}{K_{ia} K_b} + \frac{K_q \cdot ap}{K_{iq} K_{ia} K_p} + \frac{pq}{K_p K_{iq}} + \frac{K_a bq}{K_{ia} K_b K_{iq}} + \frac{abp}{K_{ia} K_b K_{ip}} + \frac{bpq}{K_{ib} K_p K_{iq}}} \left(1 - \frac{\frac{pq}{ab}}{K_{eq}} \right)$$

Both numerator and denominator of β are functions of the concentrations of the reactants, with Michaelis and inhibition constants as coefficients. The fact that a reaction operates near equilibrium implies that the levels of its substrates and products are strongly correlated, although not necessarily proportional (that would mean pseudo-equilibrium). Thus, all the concentration terms in the denominator are correlated and, if their sum is in the order of one or higher, the whole denominator becomes correlated with the numerator (for ping pong reactions this is always true since there is no constant term in the denominator). That renders the overall value of β comparatively insensitive to changes in the metabolites, which provides a mechanistic justification for our observations. With some reactions for which *in vitro*-determined Km values are available from literature (though relative to other strains/species) we could estimate that changes in the predicted value of β were indeed much smaller than

changes in the respective reactant concentrations (results not shown). In addition, even if there are significant changes in β, as long as the maximal displacement from equilibrium is modest (e.g. Q/Keq>0.5), the resulting variations in the slope of the Q/v relation may be nearly indistinguishable from experimental noise in the data. If so, from a practical point of view, a linear function still constitutes an adequate approximation with good predictive abilities.

Although this situation corresponds to a radical simplification of the more complex mechanistic rate equations, its prevalence among the reactions analyzed suggests that it is a recurrent feature of metabolic reaction networks, which is worth exploiting. Therefore, expressing v as a linear function of Q may be regarded as a physiologically relevant approximative kinetic format, which we shall refer to as "Q-linear kinetics". In its simplest form, used here, the two parameters of "Q-linear kinetics" are βV_{max} and K_{eq}. But in a broader sense, protein levels should not be treated as constants since they are known to change between growth conditions or in response to stimuli. By definition, $V_{max}=E.k_{cat}$ (where E is the enzyme level), so a more general rate expression can be written as a function of the parameters βk_{cat} and K_{eq}:

$$v = \beta V_{\max} . \left(1 - \frac{Q}{K_{eq}}\right) = E . \beta k_{cat} . \left(1 - \frac{Q}{K_{eq}}\right) \qquad \text{(equation 6.4)}$$

In the foreseeable future it will be possible to provide absolute enzyme levels to the model, via targeted proteomics measurement techniques [12, 136]. Ultimately, successful integration of models of the metabolic network with models for upstream regulation (e.g. transcription, translation, protein degradation) should allow protein levels to be treated as dependent model variables, rather than independent variables.

Note that we purposefully avoided the term "mass-action" to prevent confusion with the generalized mass-action kinetic format, since the resulting kinetic expressions share some similarity but are not the same. For example, the mass action kinetic expression for a uni uni reaction (A↔B) is:

$$v = k_1 . [A] . \left(1 - \frac{k_{-1} . [B]}{k_1 . [A]}\right) = k_1 . [A] . \left(1 - \frac{Q}{K_{eq}}\right) \qquad \text{(equation 6.5)}$$

The additional linear dependence on the substrate concentration ([A]) is crucial because as a consequence this kinetic expression cannot be fitted to the *in vivo* data.

It is also interesting to observe that an expression for a linear relation between v and Q can be arrived at even without kinetic or mechanistic considerations. From a thermodynamic point of view, the rate of a process close to equilibrium may be described as proportional to the free-energy difference, by so-called flow-force relations [75, 196]:

$$v = L . \left(-\Delta_r G\right) = L . \left(-RT\right) . \ln\left(\frac{Q}{K_{eq}}\right) \qquad \text{(equation 6.6)}$$

The proportionality factor L is referred to as phenomenological coefficient [132, 133]. Around equilibrium $Q/K_{eq}-1\approx 0$ and the truncated linear expansion of the logarithm results in a relation analogous to the "Q-linear kinetics" format:

$$v = L \cdot (-RT) \cdot \ln\left(1+\left(\frac{Q}{K_{eq}}-1\right)\right) = L \cdot RT \cdot \left(1-\frac{Q}{K_{eq}}\right) \qquad \text{(equation 6.7)}$$

In practice, because the reactions are close to equilibrium, plotting Q or ln(Q) as function of v results in very similar profiles, which can both be reasonably approximated by linear functions (not shown).

In fact, the occurrence of regions of linearity between (the logarithm of) Q and v has been described before, mainly in energy-transduction systems (e.g. [87]). It has also been shown, starting from mechanistic rate equations, that near-linear behavior could arise around the inflection points of predicted sigmoid-shaped relations between ΔG and reaction rate [151, 174]. However, the predictions of a sigmoid-shaped profile are the consequence of the assumption that either the substrate concentration or the product concentration [151] or their sum [174] remains constant (implying dA/dv=0, or dP/dv=0, or dA/dv=–dP/dv). Considering that the concentration of nearly every metabolite varies by more than 2-fold over the range of conditions investigated here (the median fold-change was 5-fold), that might be an acceptable assumption for a very small number of near-homeostatic reactant pairs (e.g. ATP/ADP), but does not hold for the vast majority of metabolites. In addition, what we observe in near-equilibrium reactions is that substrate(s) and product(s) change in consonance and in the same direction (implying dA/dv≈dP/dv), not the opposite. Thus, we conclude that it is the correlations in reactant levels that lead to a nearly constant β and, hence, to the ability to apply "Q-linear kinetics".

The main advantage of the "Q-linear kinetics" format is that it involves only two parameters which, as we shall demonstrate, can be easily determined from available *in vivo* experimental data. Another key advantage is that the parameters retain a mechanistic meaning, thus facilitating the interpretation of results, which is not always clear-cut with approximative formats. In fact, it can be argued that this 2-parameter format provides the optimal degree of complexity to match the (linear) kinetic behavior displayed *in vivo* by the reactions in this category. Employing the complete mechanistic rate expression, with the associated affinity parameters, would be of little use. The same way that the overall value of β is relatively insensitive to combined changes in the reactants, it will also be insensitive to combined changes in the parameters (Table 6.1). Thus, the values of individual affinity parameters would not be observable from the *in vivo* data. A concrete example of this non-identifiability problem is given in Figure 6.7, for the "best-case-scenario" of the uni uni reaction PGI. Even with 32 data points, over a 63-fold range of fluxes, and an average standard deviation of 7% in the metabolite measurements, the two Km values still cannot be separately identified (as indicated by the sum-of-square-residuals (SSR) landscape).

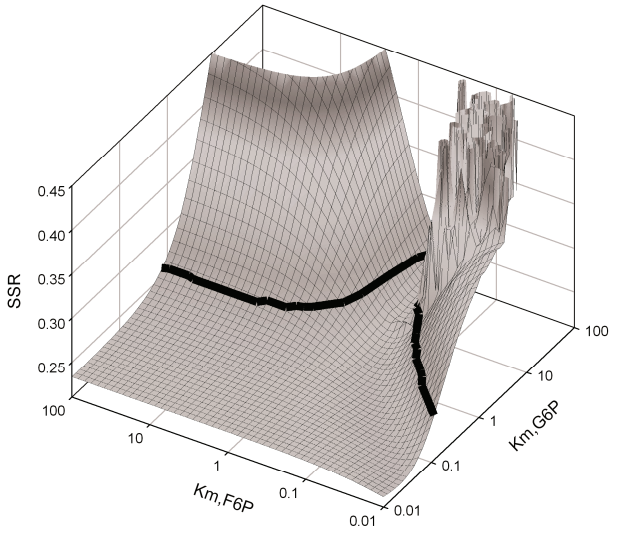


Figure 6.7: Example of non-identifiability of affinity parameters from *in vivo* data, in the case of the uni uni reaction PGI. The landscape represents the sum of square residuals (SSR) between measurements and model prediction (mechanistic rate equation), for combinations of Km values for substrate and product (Vmax and Keq are set free). The bold solid line indicates the maximum threshold SSR corresponding to the 95% confidence interval in the parameter values (the area below it indicates statistically acceptable Km values). Note that the X and Y axis are in log scale. Despite having 32 data points, over a 63-fold range of fluxes, and an average standard deviation of 7% in the metabolite measurements, the two Km values still cannot be separately identified.

In contrast, further reducing complexity would mean replacing "Q-linear kinetics" with pseudo-equilibrium. However, considering that the measured Q dropped, on average, by more than half between low and high fluxes, wrongfully applying the pseudo-equilibrium assumption would result in substantial bias (>2-fold) in the prediction of metabolite concentrations. Given the high degree of connectivity of the metabolic pathways, these errors would propagate along the network and their effect can be cumulative. As an example, consider the series of reactions in central and lower glycolysis leading from FBP to PEP: FBA, TPI, GAPDH, PGK, GPM and ENO. Since they are all close to equilibrium, one might be tempted to treat the entire set as one pseudo-equilibrium pool. However, at high fluxes some of the reactions reach displacements from equilibrium of 0.3-0.6. The bias introduced by the pseudo-equilibrium assumption at each reaction accumulates, so the prediction of PEP concentrations from FBP, or vice-versa, would reach estimation errors of over 10-fold (overestimation of PEP or underestimation of FBP). These two metabolites are respectively the substrate and allosteric activator of PYK, a key regulatory enzyme, so the prediction of flux in that reaction would be severely compromised by simplifications made elsewhere in the network. In a similar manner, assuming pseudo-equilibrium for the reactions in the non-oxidative branch of the PPP or among the hexose-phosphates would lead to overestimation of products or underestimation of substrates by as much as 2- to 3-fold, respectively. These examples illustrate how excessive simplification of near-equilibrium reactions can spread and amplify along the network and, by affecting the prediction of substrates and allosteric effectors of far-from-equilibrium reactions, can affect the prediction of the overall system behavior.

Apparent in vivo Keq data and their potential for wider applicability in thermodynamic analyses

It follows from the considerations above that it is straightforward to derive the *in vivo* parameter values of Keq and βV_{max} for the pseudo- and near-equilibrium reactions from the linear regression of Q versus v. This is a very exceptional outcome because it provides, simultaneously, fully *in vivo*-derived mathematical descriptions for all the reactions in these two categories, which can be readily incorporated into kinetic models. An overview of the estimated *in vivo* parameter values and the comparison with literature Keq data is presented in Table 6.2. Furthermore, to our knowledge, this is the first report in the literature of systematic estimation of apparent *in vivo* K_{eq} values.

Table 6.2: Overview of the *in vivo*–derived parameter values for pseudo- and near-equilibrium reactions and comparison with literature Keq data. The p-value relates to the probability of the slope of the weighted linear regression of Q versus v being zero (see text for details). When relevant, the units of concentration are mol/l.

Data source:		This study				Literature
Enzyme and EC number		p-value	Class	Apparent *in vivo* Keq	β.Vmax (μmol/g_{DW}/s)	Keq
TK(2)	*2.2.1.1b*	5.7 x10^{-1}	Pseudo-equilibrium	5.91 (±0.42)		19 (±11)
ADK	2.7.4.3	4.2 x10^{-1}		0.962 (±0.017)		1.17 (±0.33)
TA	2.2.1.2	2.2 x10^{-1}		0.1178 (±0.0058)		0.75 (±0.35)
ACO	4.2.1.3	8.7 x10^{-2}		0.0956 (±0.0029)		0.063 (±0.016)
PMI	*5.3.1.8*	3.1 x10^{-2}		1.183 (±0.013)	-	0.8 (±0.22)
GAPDH &PGK	1.2.1.12 & 2.7.2.3	2.9 x10^{-2}		113.3 (±5.9)		870 (±280)
RPE	5.1.3.1	1.0 x10^{-2}		3.102 (±0.059)		1.71 (±0.8)
TPP	3.1.3.12	7.3 x10^{-3}		22.71 (±0.93)		50
TPI	5.3.1.1	5.6 x10^{-3}		0.0391 (±0.0021)		0.0478 (±0.0054)
Fum	4.2.1.2	6.0 x10^{-5}	Near-equilibrium	5.18 (±0.14)	2.99 (±0.63)	4.29 (±0.66)
RPI	5.3.1.6	2.7 x10^{-5}		2.12 (±0.14)	0.303 (±0.063)	3.1 (±1.3)
TK(1)	2.2.1.1a	1.4 x10^{-5}		0.35 (±0.031)	0.227 (±0.047)	1.6 (±0.69)
FBA	4.1.2.13	1.2 x10^{-6}		1.038 (±0.07) x10^{-5}	3.68 (±0.64)	1.01 (±0.29) x10^{-4}
PGM	5.4.2.2	5.1 x10^{-7}		0.063 (±0.0035)	0.199 (±0.033)	0.0576 (±0.0034)
PMM	5.4.2.8	2.9 x10^{-7}		0.0872 (±0.0026)	0.178 (±0.028)	0.39 (±0.53)
ADH	1.1.1.1	1.2 x10^{-7}		4370 (±180)	3.25 (±0.25)	10600 (±3400)
ATA	2.6.1.2	5.9 x10^{-8}		0.455 (±0.035)	0.0552 (±0.0068)	1.4 (±0.47)
Eno	4.2.1.11	2.7 x10^{-18}		4.011 (±0.093)	3.75 (±0.19)	4.08 (±0.73)
PGI	5.3.1.9	1.6 x10^{-21}		0.2586 (±0.0028)	4.07 (±0.17)	0.319 (±0.078)
GPM	5.4.2.1	1.1 x10^{-21}		0.11576 (±0.00084)	7.11 (±0.27)	0.0921 (±0.0044)

Having obtained *in vivo* values of these parameters, it is worth evaluating how the results obtained relate with available *in vitro*-derived data. In the case of βV_{max} a realistic comparison is in fact not possible because, for the same strain and growth conditions, Vmax data is only available for a few of the enzymes and the affinity parameters necessary to compute β are not available at all. As for Keq, which are not strain-specific or condition-dependent, we can compare the results with the respective *in vitro*-derived literature values.

We find that for many reactions the estimated apparent *in vivo* K_{eq} are very close to the respective *in vitro* values, notably for ACO, GPM, FMH, PGM, ENO, TPI, ADK, and PGI, all differing by less than 25%. However, there are also some significant differences. For RPE and PMI, the apparent *in vivo* K_{eq} is higher than the *in vitro* values, so in a thermodynamic analysis based on the *in vitro* Keq these reactions would appear to be infeasible. The values for RPI, APT, PMM and GAPDH+PGK are lower than the experimental Keq data, but comparable to values derived from ΔG data of Alberty and Jankowski et al (not shown), highlighting the discrepancies between literature sources. The largest differences are found with ADH, TK2, TK1, TAL and FBA, for which the apparent *in vivo* K_{eq} are below the lowest values from any literature source. Consequently, a kinetic analysis based on the respective *in vitro* Keq would greatly overestimate their *in vivo* rate. Interestingly, these Keq relate to the reactions with the most disperse Keq information while, in contrast, the reactions for which there is least uncertainty in the *in vitro* data (e.g. ENO, TPI, PGM, FMH and GPM) are also those with best agreement between *in vivo* and *in vitro* values. There are several possible reasons for the differences observed, namely the potential for inaccuracies in Keq data (particularly when data is scarce, such as with the PPP reactions) and the possibility of experimental error in the metabolite measurements. However, it is also essential to note that *in vivo* and *in vitro* Keq need not always be the same. The equilibrium between reactants *in vivo* may differ due to several factors, some of which are difficult to take into account because the intracellular conditions are not sufficiently characterized (e.g. pH, ionic strength, specific cell volume, concentration of specific ions [150], cosolvent effects [14], nonspecific interactions with cellular components [119]). On the other hand, there are many reasons why metabolite measurements may not provide the true *in vivo* concentrations (e.g. compartmentation [33], protein-binding [26], non-homogeneity [90, 186]). Taking these aspects into account, while the many similarities between estimated apparent *in vivo* values and *in vitro*–derived data give credibility to the approach, the observation of some differences is not at all unexpected.

Strictly speaking, the apparent *in vivo* Keq values cannot be considered true Keq because it can never be guaranteed that metabolomics data reflects true concentrations, but from a practical point of view they may be exceptionally useful, because *in vivo*-derived Keq data provide an ideal, compatible reference. However, to be used as reference, the values must be approximately constant. A crucial question is therefore to what extent the values obtained are valid under other growth conditions or even in different cells/organisms. This matter cannot be fully addressed yet for lack of data but to get a first impression we compared our results with recent metabolite data obtained for *E. coli* [166]. It seems reasonable to assume that at the very low specific growth rate investigated in that study (0.1 h^{-1}, approximately 13% of μ_{max}), the measured Q should be close to the *in vivo* Keq. Surprisingly, the reported Q values for ADK (1.01), PGI (0.27), PMI (1.27) and FMH (4.3) are all within 16% of the corresponding

apparent *in vivo* Keq we observed in *S. cerevisiae* (a fifth reported ratio is undermined by analytical artifacts uncovered later). In contrast, the differences relative to literature Keq data reached 60% (or 80% and 5-fold, if based on ΔG data of Alberty and Jankowski et al, respectively [4, 85]). In other words, it would be possible to make more precise thermodynamic considerations in *E. coli* based on the apparent *in vivo* Keq values derived from yeast than using *in vitro*–derived Keq data from literature. Admittedly, both are unicellular species grown under very similar conditions (aerobic, glucose-limited), but the fact that one is a eukaryote and the other a prokaryote is sufficient to make the similarity very remarkable. In our view, the degree of agreement between the values seems extremely promising for the validity and wider application of apparent *in vivo* Keq data. We expect that further studies applying the same *in vivo* approach with other microbes and growth conditions (to account for differences in intracellular environment) as well as different analytical pipelines (to account for analytical error) should reveal the full extent of applicability of the Keq obtained here. Looking forward, wider adoption of our approach combined with improvements in coverage of quantitative metabolomics techniques, should enable the high-throughput estimation of apparent Keq for large numbers of reactions. This would be particularly useful regarding the lesser studied reactions, for which Keq data is scarce or missing. In turn, such information will provide a breakthrough in network-wide thermodynamic analysis, for which the large uncertainties in thermodynamic data are currently a key obstacle [55, 76, 99, 118].

Finally, regarding the experimental strategy for determination of the apparent *in vivo* Keq, it is relevant to note that the approach is readily applicable to other organisms and easily scalable. As long as the flux through the metabolic network can be varied in a well-defined manner, apparent *in vivo* Keq values can be obtained simultaneously for as many (pseudo-/near-equilibrium) reactions as the coverage of the analytical platform allows. In terms of experimental design, we chose to use chemostat cultures at different dilution rates for convenience and because of the wide range of fluxes that can be achieved. This had the unexpected disadvantage that at low growth rates many metabolite concentrations approach zero, increasing the chance of analytical bias. There are several alternatives. For example, one could use different substrates or mixtures at varying ratios (e.g. glucose/ethanol [179]), which can have the advantage that the flux through certain reactions will at some point reverse, allowing the Q at zero flux (i.e. Keq) to be interpolated rather than extrapolated.

Estimation of network-wide distribution between reaction categories and implications for large-scale kinetic models

Having sorted each of the 27 reactions into one of the three categories, we can now analyze the resulting distribution, which is depicted in Figure 6.8 (panel A). The far-from-equilibrium reactions represent one quarter of the reactions, while the remaining are pseudo- or near-equilibrium. According to these results, the organization of the central metabolism of yeast is not far from the classical, "textbook" view of metabolism as sequences of reactions close to equilibrium interspersed with a few reactions far from equilibrium, at key positions of the pathways, controlling the distributions of fluxes.

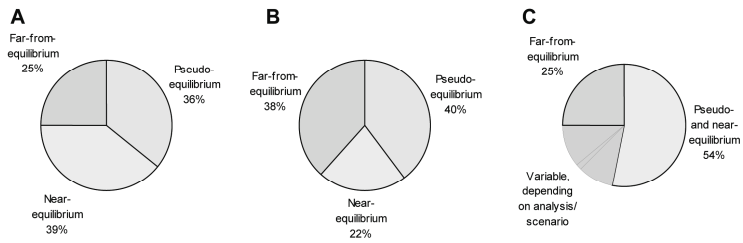


Figure 6.8: Estimations of network-wide distribution of reaction categories A: proportion of reaction categories among the reactions analyzed in this study; B: extrapolated distribution of reaction categories via normalization with the relative frequency of each EC class in the metabolic network of yeast; C: for comparison, the range of estimated distributions based on two thermodynamic analyses of genome-scale models of *E. coli* [55, 76] (including 3 different scenarios in the case of Fleming et al).

However, this distribution of the three categories is based on a total number of reactions that is relatively small, if compared to the size of the whole metabolic network. It could also be biased because the subset of reactions analyzed is drawn entirely from central carbon metabolism. So what lessons can be drawn having in mind the ultimate goal of constructing genome-scale kinetic models? Although making such extrapolations is inherently risky, we attempted to answer this question by looking for relations between the distribution among reaction categories and the EC classes of the respective enzymes. The idea behind this was that certain types of biochemical reactions, for which we used EC class as proxy, might be more likely to operate close to or far from equilibrium. Indeed, some interesting patterns emerged. For example, our dataset contains eight reactions of EC class 5 (isomerases), and not a single one of them operates far from equilibrium. Near-equilibrium reactions are a minority in EC class 2 (transferases), where the distribution (pseudo/near/far-from-equilibrium) is 40%:20%:40%, but a majority among EC class 4 (lyases), where the distribution is 20%:60%:20. Unfortunately, in our dataset there are few reactions of EC class 1 (oxidoreductases) and 3 (hydrolases) and no reactions of EC class 6 (ligases). In this aspect, the subset of reactions analyzed here notably differs from the estimated relative abundance of each class in the metabolism of yeast and other organisms, where EC classes 1, 2 and 3 predominate, together representing 80-90% of all reactions. We therefore attempted to estimate a network-wide distribution of classes by taking into account the percentage of each category per EC class and the relative abundance of each EC class in metabolism (according to the Biocyc database [91]). The resulting distribution was 40%:22%:38% (Figure 6.8, panel B). Thus, based on this estimation and the observed distribution in central metabolism, we conclude that an approximately even distribution between the three reaction classes constitutes a reasonable approximation of the overall distribution in the metabolic network of *S. cerevisiae*.

We then looked into the literature for reported network-wide thermodynamic analyses. Two recent studies were found [55, 76], reporting thermodynamic analyses based on genome-scale models of *E. coli* containing 869 and 2077 reactions, respectively. Henry et al estimate a proportion of reactions operating far from equilibrium of 38%, while Fleming et al obtain percentages of 25, 36 or 47% depending on assumptions regarding physiological ranges of

metabolite levels (Figure 6.8, panel C). These proportions are very close to our estimates, which is quite remarkable considering they were arrived at by different methodologies and refer to different organisms. Taken together, they lend credibility to our estimated distributions, setting the likely fraction of far-from-equilibrium reactions at around 30-40%. They also suggest that the proportions may be very similar among different organisms, perhaps reflecting a conserved mode of organization and regulation of metabolism, favored under certain growth conditions or by evolutionary pressures. In the foreseeable future, advances in quantitative metabolomics should allow our approach to be applied to much larger sections of the metabolic network, which will brings us closer to the true distribution and provide further insights into the thermodynamic organization of metabolism.

Interestingly, the more refined estimates put the proportion of far-from-equilibrium reactions close to 40%, which is somewhat higher than observed within the dataset analyzed in this study. This is mostly a consequence of the relative abundance of isomerisation reactions in central metabolism, all of which are observed to operate close to equilibrium. Retrospectively, it is possible that because the reactions in central metabolism were, historically, the most experimentally accessible for kinetic studies *in vivo*, the relative prominence of reactions close to equilibrium would have played a role in shaping the classical notion of metabolism as involving only a small minority of "pace-maker" reactions operating far from equilibrium.

A proportion of 30-40% of far-from-equilibrium reactions has important implications for efforts aiming at large-scale kinetic modeling. Even a fairly modest network-wide model will contain well over 100 far-from-equilibrium reactions, many of which may require very detailed descriptions. A great deal of highly informative *in vivo* data will still be needed to allow the proper estimation of the multitude of associated kinetic parameters. This observation emphasizes the need for structured approaches for submitting reactions identified as far-from-equilibrium to further rounds of model reduction, especially via approximative kinetic formats [27, 74]. It also highlights the need for development and improvement of methodologies for analysis of information content and model-based experimental design [124, 126].

On the other hand, an estimated 60-70% of reactions operating as pseudo- and near-equilibrium also confirms that there is a huge scope for applying data-driven simplifications to kinetic models. Based on our findings, the extent of simplification that can be afforded is indeed very substantial, leading to meaningful, predictive mathematical descriptions (apparent *in vivo* Keq and "Q-linear kinetics") which require few, *in vivo*-identifiable parameters. In this respect, we believe that the approach proposed here to distinguish pseudo- and near-equilibrium reactions, leading directly to their *in vivo* parameterization, will play a pivotal role in facilitating the formulation of large-scale kinetic models of metabolic systems.

CONCLUSIONS

The construction of realistic large-scale *in vivo* kinetic models can be greatly facilitated if the first stage of model formulation is guided by the *in vivo* data itself, via the reaction classification scheme proposed here. This framework achieves the double objective of a) allowing the

complexity of mathematical description of each reaction in the model to be precisely tailored to the complexity of kinetic behavior displayed *in vivo*, and b) simultaneously providing a fully *in vivo*-derived description, that can be directly incorporated into kinetic models, for reactions in two of the categories (pseudo- and near-equilibrium), which together represent an estimated 60-70% of the metabolic network. Reactions in the third category (far-from-equilibrium) may be suited for further rounds of model reduction, for example via approximative kinetic formats.

As a model formulation strategy, the reaction classification scheme is rooted in thermodynamics, conceptually simple and yet very powerful in the model simplifications it reveals. An important advantage is that it requires no *a priori* kinetic knowledge aside from estimates of Keq. Instead, it relies heavily on the ability to generate high-quality quantitative *in vivo* metabolomics and proteomics data. Although we have demonstrated its use in the central metabolism of *S. cerevisiae*, the methodology is readily scalable and, in principle, applicable to any other biological system.

The *in vivo* kinetics of near-equilibrium reactions can be conveniently described as a linear function of the reaction quotient, a kinetic format which we dubbed "Q-linear kinetics". The reason the simplification is possible can be traced back to intrinsic correlations between changes in reactant levels under *in vivo* conditions. Considering the prevalence of near-equilibrium reactions in the metabolic network, we expect this useful kinetic format to become widespread in efforts aiming at large-scale *in vivo* kinetic modeling.

The exploitation of the Q(v) plots allowed for the first time the systematic estimation of apparent *in vivo* Keq values. Preliminary results indicate the data are inter-changeable between different organisms and constitute a better reference for precise thermodynamic considerations based on *in vivo* metabolite data than *in vitro*-derived Keq values. The full extent of applicability of *in vivo* Keq is not yet known, but the application of our *in vivo* approach to simultaneously estimate Keq for large numbers of reactions is limited only by the analytical coverage of the network.

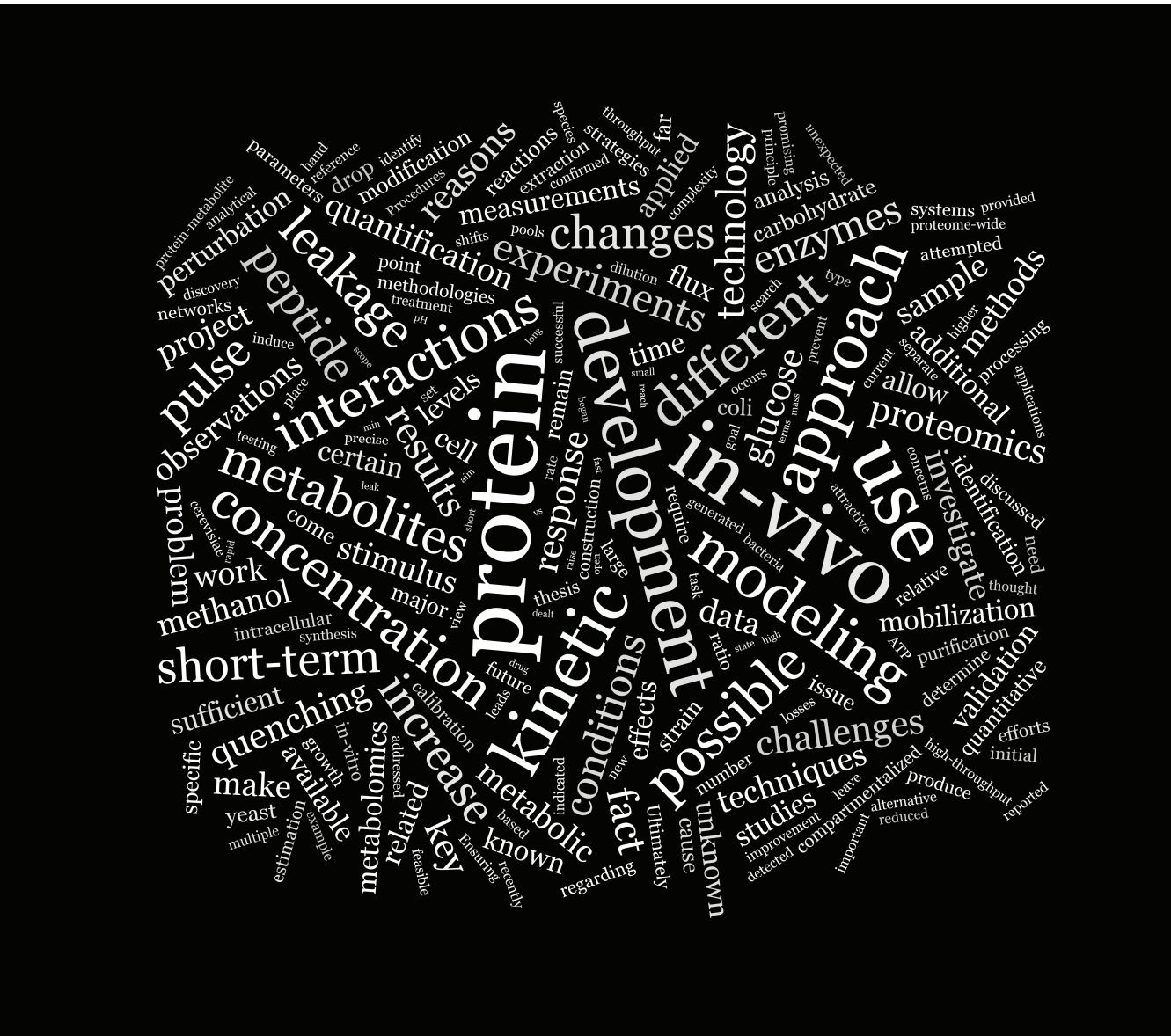
ACKNOWLEDGEMENTS

The authors are grateful to Peter Verheijen, Aljoscha Wahl and Domenico Bellomo for critical discussions and suggestions, Hans Westerhoff for pointing out the connection between "Q-linear kinetics" and flow-force relations, Karen van Eunen, Chiara Cippolina, Rick Orij and Gertien Smits for sharing unpublished data on cell volume and intracellular pH, Lodewijk de Jonge and Rutger Douma for advice on the rapid filtration method, Reza Seifar, Zhen Zeng and Max Zomerdijk for expert analytical work, and Dirk Geerts and Rob Kerste for knowledgeable support in the fermentation lab.

This work was funded by SenterNovem through the IOP Genomics initiative (project IGE3006A) and carried out within the research program of Kluyver Centre for Genomics of Industrial Fermentation, which is part of NGI/NWO.

Chapter 7

Reflections & Outlook



Rather than repeating the conclusions and recommendations of the previous chapters, regarding each of the challenges which were dealt with, I chose to use this last chapter to discuss the challenges which were not tackled, some of the many observations and results which are not reported in this thesis and certain developments which have taken place since the initial work was performed. Some views on the strategies which are most likely to bring us closer to realizing the ambition of *in vivo* kinetic modeling will also be discussed.

TWO OTHER KEY CHALLENGES FOR *IN VIVO* KINETIC MODELING

In the general introduction to this thesis I began by outlining the four key challenges which would be addressed in the subsequent chapters. Those are indeed crucial hurdles in the path towards quantitative analysis of *in vivo* kinetics, but in fact there may well be several others. I will make no attempt to come up with an extensive "wish list" but there are two issues, not dealt with in this project, which I think are important to acknowledge and discuss, namely the need for quantitative proteomics and the identification of *in vivo* metabolite-enzyme interactions.

Quantitative proteomics

In addition to metabolite concentrations, kinetic modeling requires data on the intracellular protein levels. For the time being, these can be provided via protein measurements. Ultimately, estimation of protein levels will be supplied by models accounting for the genetic regulation of protein abundance, but the construction of those models will also require protein measurements.

In principle, proteome-wide quantification is already feasible using current MS-based peptide mass fingerprinting techniques. Although what is often referred to as quantitative proteomics is in fact *relative* quantification based on isotope dilution strategies (e.g. SILAC, ICAT, iTRAQ), *absolute* quantification is also possible as long as reference proteins or "proteotypic peptides" of known concentration are available for calibration [62, 136]. Therein lies a key challenge, since the availability of the calibration standards requires, for every strain/species, either proteome-wide purification of the individual enzymes or, more likely, the synthesis of the respective (isotopically-labeled) proteotypic peptides. One approach that facilitates peptide synthesis is to use so-called QconCAT gene constructs, which allow the biosynthesis of multiple peptides in parallel via a concatenated peptide [12, 140]. Proteome-wide coverage will still require a large number of constructs but since the technology has applications in multiple research fields and is already being commercialized, the availability of large peptide calibration sets for model organisms appears possible in the near future. On the other hand, a more fundamental problem of peptide mass fingerprinting will be harder to tackle. Each protein is determined via just a few peptides, selected mostly on the basis of their MS response properties. If an important post-translational modification occurs elsewhere in the molecule it will not be

detected. This also presents a problem for absolute quantification, since calibration peptides must be produced for the different modification variants of the target protein.

A separate aspect of accurate quantification of intracellular proteins, which so far seems to receive little attention, is the sample processing. Current protein isolation methods (e.g. sonication, bead-beating) do not necessarily achieve complete extraction. Procedures tend to be biased in relation to certain types of proteins (e.g. cytosolic vs membrane-bound) or are targeted for specific classes (e.g. phosphorylated). Furthermore, although protein translation occurs at much longer time-scales, covalent modification is a fast process. For these reasons, the same considerations that are relevant for metabolomics, in terms of quenching and extraction, also apply to proteomics if accurate quantification and preservation of post-translational modification are the goal. In principle, it is quite possible that the techniques validated for metabolomics in a certain cell type can be directly applied for proteomics. Ultimately, a uniform sampling and processing scheme could be used for all omics analyses. However, it also possible that proteins turn out to be more sensitive to degradation or chemical changes under certain treatment conditions.

Identification of metabolite-enzyme interactions

So far, *in vivo* kinetic modeling efforts have aimed only at estimating the values of parameters for known protein-metabolite interactions, namely, in respect to reactants and allosteric effectors already known from *in vitro* studies (and even that has proven difficult!). However, it is possible, even likely, that many interactions taking place *in vivo* remain unknown. Enzymologists often test the effects of compounds similar to the natural substrates and/or intermediates in other reactions of a common pathway., but less intuitive compounds are likely to go untested, as illustrated by the fact that important, yet unexpected regulators are still occasionally found (e.g. [15, 180]). At a conservative number of 1000 metabolites and 1000 enzymes, the search space of possible combinations would be 10^6. Discovering which interactions are relevant and which are not will be a colossal task, which has yet to be undertaken.

The most obvious, if cumbersome, approach is to kinetically (re-)characterize individual purified enzymes against comprehensive libraries of metabolites. Method development can be facilitated by the use of MS-based analytics [60, 61], which circumvent the need for coupled reactions (e.g. for NAD(P)H-based spectrophotometry) and allow assaying in both directions. This approach has already been demonstrated for the discovery of new enzyme activities [154]. Development in micro-/nano-fluidics may contribute to achieving higher throughputs and reducing reagent usage. However, some of the major challenges include the synthesis/purification of the metabolite libraries, the purification of enzymes and the development and validation of methods for sufficiently high-throughput analysis.

A very promising approach would be to make use of the high-throughput screening technologies being developed for drug discovery and mapping of protein-protein and protein-ligand interactions. Conceivably, protein-microarray or small-molecule-microarray techniques could be employed to globally assess protein-metabolite interactions [122, 142]. Some specific challenges regard the manufacture of the chips and ensuring sensitivity and specificity of the

interactions detected. Nevertheless, since this type of technology has potential applications in other fields, including drug screening and toxicology, it is likely to benefit from increased interest and rapid technology development.

On the other hand, the "*in vivo* vs *in vitro* dilemma" may also apply to the identification of protein-metabolite interactions, since purification and immobilization can affect protein conformation, as well as the structure of enzyme complexes. If feasible, an *in vivo* discovery approach would be preferable. However, if the information content of *in vivo* experiments is often not even sufficient for reliable estimation of parameter values for known interactions, then it is unlikely that it can be used to identify the unknown interactions. This might change, however, when advances in high-throughput cultivation, fluxomics, metabolomics and proteomics reach a stage where it is feasible to analyze very large numbers of closely related strains (e.g. over-expression and deletion mutants) under many growth conditions. A key challenge, besides the technology development, would be to design experiments such that the effects of highly correlated metabolites can be resolved [126].

TO LEAK OR NOT TO LEAK

Although it came to represent a substantial part of this thesis, the improvement of metabolomics sample treatment methodologies was not initially a goal of the project (the available methods were thought adequate). What began as a simple task of testing milder extraction techniques (thought necessary for future measurements of labile metabolites, such as the CoA esthers) developed into a search for the reasons behind the lack of reproducibility of metabolite data and later, once suspicions regarding changes during quenching were confirmed, into a full (and long) push to find a solution to the leakage problem, and ensure that subsequent studies would be based on solid data. It would be fitting at this point to proclaim the issue closed but, as far as leakage is concerned, there are reasons to think that the story is far from finished.

The work described in Chapter 2 offers rather convincing evidence for the occurrence of extensive leakage of metabolites from yeast cells as well as a relatively simple remedy for the problem: increase the methanol concentration. However, around the same time that work was published, another article was published investigating the same problem but reaching the opposite conclusions [17]. According to that study, leakage using the conventional method (60% methanol) was minimal and increasing the methanol concentration actually increased losses slightly. The article was from the only other group that has so far investigated this issue quantitatively and their methodology seems sound, so the unavoidable question is: why the difference in outcome? At first sight, there are only minor differences in the experimental procedures: the strain used (ATCC 32167), the cultivation mode (shake-flasks), the temperature profile upon sampling (quenching solution initially at -58°C) and the fact that whole-broth, quenching supernatants and extracellular filtrate samples were not extracted prior to concentration (leaving the possibility of enzymatic conversion). The fact that, based on our current knowledge, none of these is sufficient to explain the difference in results is a reminder that we do not yet fully understand the causes of this phenomenon.

Further work carried out within our own lab has also provided additional puzzling observations. Investigations in *E. coli* revealed massive leakage upon conventional quenching [166]. Interestingly, increasing the methanol concentration reduced leakage, as observed in yeast, but the improvement was not sufficient to prevent the losses entirely. Leakage was also observed with *Penicillium chrysogenum* but, unlike with *S. cerevisiae* and *E. coli*, increasing the methanol concentration was found to increase losses (Lodewijk de Jonge and Rutger Douma, manuscript submitted). The reason for the different reaction to methanol is unknown. Even more recently, validation work with the yeast *Pichia pastoris* has shown the absence of any significant leakage, regardless of the methanol concentration used, between 40 and 100% (Marc Carnicer, unpublished results). The reason for the apparent superior robustness is unknown, although it has been speculated that it could be related to *P. pastoris* having a thicker cell wall.

So it seems clear that leakage extent is species-dependent. Additionally, the differences between studies with *S. cerevisiae* point at the possibility of strain dependency and perhaps even growth condition dependency. If confirmed, that would be a truly nightmarish scenario because of the validation burden that it would impose on each new set of experiments.

On the other hand, there are also reasons to be optimistic. We have recently performed a set of "make or break" trial experiments in *E. coli* and were able to identify very promising alternatives towards preventing leakage. Confirmation of the preliminary results and development of the interesting leads could produce a leakage-free quenching method for *E. coli*. Considering that gram-negative bacteria appear to be the worse-case-scenario in terms of leakage [16], such a development could well open the door for a universal quenching technique for microbial metabolomics.

A FAIL-PROOF TRICK TO BYPASS COMPARTMENTATION

After the apparently successful approach described in Chapter 4 to determine the cytosolic free NAD/NADH, an interesting prospect was to apply the same strategy for the mitochondrial ratio, as well as for each of the NADP/NADPH ratios. However, a quick search for suitable indicator reactions failed to produce good leads and, because it was not a major objective, the idea was postponed. Certainly, it is possible that in the future suitable indicators may be found for these cofactor ratios, and that reporter proteins are designed for specific key intermediates, but that still leaves out all the other metabolites which are also compartmentalized. And considering the difficulties encountered in quenching whole cells, it is unclear when or whether it will be possible to effectively combine cell fractionation with prevention of enzyme activity, at least into a methodology practical enough to be used in high throughput. Most likely, compartmentation will remain a fundamental hindrance in the *in vivo* study of metabolic systems in eukaryotes for years to come. Efforts to overcome this limitation are certainly worthy, from the point of view of enabling the study of metabolism in eukaryotes, but from the point of view of developing tools and testing approaches for *in vivo* kinetic modeling the question that should be asked is: why work with compartmentalized systems when structurally simpler systems exist? The absence of intracellular transport, parallel reactions taking place in

different compartments and multiple separate pools of each metabolite should make the metabolic networks of prokaryotes much more accessible to detailed kinetic analysis, thus maximizing the chances for success and progress. Several bacteria such as *E. coli*, *B. subtilis* and *Pseudomonas* species are at least sufficiently characterized and amenable to controlled experimentation to allow the development and application of new methodologies. Approaches and concepts that prove successful in bacteria could then be transferred and attempted in fungi and higher cells, once the necessary analytical technologies are available. But until then, it seems preferable to leave out as many layers of unnecessary complexity as possible.

SHORT-TERM STIMULUS RESPONSE EXPERIMENTS: QUO VADIS?

A very substantial part of the time of this project was dedicated to developing improved short-term stimulus response strategies. The main objective was to expand the repertoire of stimuli that could be applied, in order to increase the wealth of data that could be generated for modeling purposes. The reason that the results of these efforts are not featured in this thesis is that the majority was ultimately unsuccessful. Many of the resulting observations may never end up in widely available publications, but since they raised some awareness regarding the limitations in the use of short-term stimulus response techniques it is worth discussing some of them briefly here.

The use of short-term stimulus response experiments has been advocated since the early 1990s as a way to investigate *in vivo* enzyme kinetics while circumventing the need to determine protein levels [30, 128, 171, 190]. The underlying principle is that within a short window of observation protein levels can be assumed to be constant, since transcription and translation are comparatively slow processes. Some caveats to this approach were always known, namely the fact that changes in covalent protein modification may also occur within seconds and, thus, cannot be excluded, and the fact that protein measurements will ultimately be needed if the model derived from data at a certain reference state is to be applied under other conditions, where protein levels may differ. In time, some additional pitfalls have emerged.

One issue relates to the estimation of fluxes in very short time scales, if the changes in extracellular concentrations (and off-gas) are too small (relative to measurement error), if the sampling rate is not fast enough and/or if the distribution of fluxes transiently shifts towards unknown pools (precluding proper balancing). For example, in the MSc projects of Stefan Jol and Mehmet Ucisik we were unable to adequately observe the changes in ethanol and glycerol production upon glucose pulses under anaerobic conditions because of the high concentrations already present extracellularly. However, it has already been shown that dynamic ^{13}C-based flux analysis can be used towards overcoming this type of problem [192].

Another key problem is related with the risk of sudden environmental changes inducing unexpected or undesirable side-effects. For example, it is now known that a glucose pulse to a glucose-limited yeast culture simultaneously induces a drop in the adenine pool [96], a drop in

pH [95] and the rapid mobilization of storage carbohydrates [2]. A change in pH could cause changes in the very kinetic parameters that we aim to determine, while the mobilization the carbohydrates "dumps" additional glucose into glycolysis. To investigate whether carbohydrate mobilization could be avoided we attempted milder glucose perturbations, namely pulses of lower concentrations and short-term (15 min) dilution rate shifts (5, 10, 20, 30 and 50% increase from $D=0.1h^{-1}$). We found that even a dilution rate up-shift as low as 10% for as little as 15 min is sufficient to cause mobilization. This, in effect, precludes the use of short-term pulses/steps/shifts of different intensities to change glycolytic flux in a precise, well-controlled manner.

As a side note on the subject of unexpected side-effects, during the MSc project of Tjerko Kamminga we found that, in contrast to a 5 or 10mM pulse, a 2.5 mM glucose pulse does not cause a sudden drop in ATP. The reasons for the difference remain unknown, but it was speculated that it could be related with a lesser drop in the NAD/NADH ratio. Curiously, "buffering" the ATP concentration throughout a glucose pulse had been the goal of Tjerko's project, although the arginine kinase-based system used for this purpose actually failed to work as expected (phosphoarginine could not be detected intracellularly in any conditions). In fact, our results contradicted in several aspects the claims of the original study reporting arginine kinase expression in yeast [34], namely by demonstrating that a) the *in vivo* conditions in *S. cerevisiae* do not favor thermodynamically the accumulation of phosphoarginine, and b) switching off the feed of a glucose-limited culture in fact does not cause an immediate drop in ATP, as reported.

But there were also some promising results, coming from attempts to diversify perturbation strategies beyond glucose pulses. Trials with alternative perturbation agents, carried out partly in the MSc project of Nathalie Bleijie, indicated that there is much more scope in alternative pulse agents than previously thought. Pulses of fructose, mannose, pyruvate, acetate, formate, octanol and benzyl acetone all produced observable short-term responses. This opens the prospect, for example, of perturbing carbon flux and redox state separately. Unfortunately, most of these perturbations also seem to induce carbohydrate mobilization, which will make it difficult to distinguish the effects of the original perturbation from the "secondary" effects of additional internal glucose supply. Nevertheless, they may provide attractive approaches in growth conditions/strains/species where carbohydrate mobilization is not an issue.

So what is the scope for the continued use of short-term stimulus response techniques in the investigation of *in vivo* kinetics? Besides the considerations above, some of which raise concerns of additional disadvantages in this approach, the major development affecting its attractiveness will probably come from elsewhere. As proteomics technology matures and eventually becomes accessible to an increasing number of laboratories, it will undermine the key advantage of short-term stimulus response experiments. Steady-state experiments (e.g. different growth conditions, different mutants) will then become comparatively attractive, due to the relative ease with which intracellular fluxes can be calculated. Short-term stimulus experiments may remain useful for the investigation of specific phenomena, particularly when a time dimension is of interest (e.g. adaptation, sensing, signal transduction), but in the systematic analysis of *in vivo* kinetics of metabolic networks, I anticipate they will play only a secondary role.

OUTLOOK: POSITIVE

Although a certain level of personal dissatisfaction is unavoidable when a task is left unfinished, one cannot help but feel optimistic that the day is approaching when the construction of large, predictive *in vivo* kinetic models of metabolic networks will be possible. After all, the key pieces are slowly coming together:

Metabolomics and proteomics have advanced remarkably in recent years and will likely continue to expand our ability to measure the components of metabolic networks more comprehensively and at higher throughput. Ensuring accuracy and precision will remain major concerns, but can surely be addressed via community efforts towards critical sample treatment validation and standardization of analytical methods (as discussed in **Chapters 2, 3 and 5**).

Furthermore, nanofluidics and protein and small molecule microarray technologies may soon permit the systematic identification of interactions between metabolites and enzymes. Conceivably, a hybrid *in vitro/in vivo* approach would consist of initial high-troughput identification of leads *in vitro* followed by validation and quantification of the kinetic parameters *in vivo*.

Finally, as the technologies allow more data to be generated, also developments in the theoretical and computational tools at our disposal should allow us to make better use of the data to understand and model metabolic systems. In particular, data-driven approaches such as the procedure outlined in **Chapter 6** should allow us to identify in a precise manner where the complexity of kinetic models can be reduced, making them both easier to grasp and more amenable to full *in vivo*-based parameterization.

References

1. Abe F, Horikoshi K. Analysis of intracellular pH in the yeast Saccharomyces cerevisiae under elevated hydrostatic pressure: a study in baro- (piezo-) physiology. Extremophiles 1998; 2:223-228.

2. Aboka FO, Heijnen JJ, Van Winden WA. Dynamic C-13-tracer study of storage carbohydrate pools in aerobic glucose-limited Saccharomyces cerevisiae confirms a rapid steady-state turnover and fast mobilization during a modest stepup in the glucose uptake rate. Fems Yeast Research 2009; 9:191-201.

3. Alberty R.A. Thermodynamics of Biochemical Reactions. Wiley-Interscience, Hoboken, New Jersey, USA; 2003.

4. Alberty R.A. BasicBiochemData3. http://library.wolfram.com/infocenter/MathSource/5704; 2005.

5. Alberty R.A. Biochemical Thermodynamics. John Wiley & Sons; 2006.

6. Baganz F, Hayes A, Farquhar R, Butler PR, Gardner DCJ, Oliver SG. Quantitative analysis of yeast gene function using competition experiments in continuous culture. Yeast 1998; 14:1417-1427.

7. Bagnara AS, Finch LR. Quantitative Extraction and Estimation of Intracellular Nucleoside Triphosphates of Escherichia-Coli. Analytical Biochemistry 1972; 45:24-34.

8. Bakker BM, Overkamp KM, van Maris AJA, Kotter P, Luttik MAH, van Dijken JP, Pronk JT. Stoichiometry and compartmentation of NADH metabolism in Saccharomyces cerevisiae. Fems Microbiology Reviews 2001; 25:15-37.

9. Barrett CL, Kim TY, Kim HU, Palsson BO, Lee SY. Systems biology as a foundation for genome-scale synthetic biology. Current Opinion in Biotechnology 2006; 17:488-492.

10. Bent KJ, Morton AG. Amino Acid Composition of Fungi During Development in Submerged Culture. Biochemical Journal 1964; 92:260-269.

11. Berger F, Ramirez-Hernandez MH, Ziegler M. The new life of a centenarian: signalling functions of NAD(P). Trends in Biochemical Sciences 2004; 29:111-118.

12. Beynon RJ, Doherty MK, Pratt JM, Gaskell SJ. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nature Methods 2005; 2:587-589.

13. Birkemeyer C, Luedemann A, Wagner C, Erban A, Kopka J. Metabolome analysis: the potential of in vivo labeling with stable isotopes for metabolite profiling. Trends in Biotechnology 2005; 23:28-33.

14. Blacklow SC, Raines RT, Lim WA, Zamore PD, Knowles JR. Triosephosphate Isomerase Catalysis Is Diffusion Controlled - Appendix - Analysis of Triose Phosphate Equilibria in Aqueous-Solution by P-31 Nmr. Biochemistry 1988; 27:1158-1167.

15. Blazquez MA, Lagunas R, Gancedo C, Gancedo JM. Trehalose-6-Phosphate, A New Regulator of Yeast Glycolysis That Inhibits Hexokinases. Febs Letters 1993; 329:51-54.

16. Bolten CJ, Kiefer P, Letisse F, Portais JC, Wittmann C. Sampling for metabolome analysis of microorganisms. Analytical Chemistry 2007; 79:3843-3849.

17. Bolten CJ, Wittmann C. Appropriate sampling for intracellular amino acid analysis in five phylogenetically different yeasts. Biotechnology Letters 2008; 30:1993-2000.

18. Brachmann CB, Davies A, Cost GJ, Caputo E, Li JC, Hieter P, Boeke JD. Designer deletion strains derived from Saccharomyces cerevisiae S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast 1998; 14:115-132.

19. Bragger JM, Dunn RV, Daniel RM. Enzyme activity down to-100 degrees C. Biochimica et Biophysica Acta-Protein Structure and Molecular Enzymology 2000; 1480:278-282.

20. Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. Molecular Biology of the Cell 2008; 19:352-367.

21. Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FCP, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. Nature Genetics 2001; 29:365-371.

22. Brent R. A partnership between biology and engineering. Nature Biotechnology 2004; 22:1211-1214.

23. Bruggeman FJ, Westerhoff HV. The nature of systems biology. Trends in Microbiology 2007; 15:45-50.

24. Bruinenberg PM, Vandijken JP, Scheffers WA. An Enzymic Analysis of Nadph Production and Consumption in Candida-Utilis. Journal of General Microbiology 1983; 129:965-971.

25. Brun MA, Tan KT, Nakata E, Hinner MJ, Johnsson K. Semisynthetic Fluorescent Sensor Proteins Based on Self-Labeling Protein Tags. Journal of the American Chemical Society 2009; 131:5873-5884.

26. Bucher T, Brauser B, Sies H, Langguth O, Conze A, Klein F. State of Oxidation-Reduction and State of Binding in Cytosolic Nadh-System As Disclosed by Equilibration with Extracellular Lactate Pyruvate in Hemoglobin-Free Perfused Rat-Liver. European Journal of Biochemistry 1972; 27:301-317.

27. Bulik S, Grimbs S, Huthmacher C, Selbig J, Holzhutter HG. Kinetic hybrid models composed of mechanistic and simplified enzymatic rate laws - a promising method for speeding up the kinetic modelling of complex metabolic networks. Febs Journal 2009; 276:410-424.

28. Buscher JM, Czernik D, Ewald JC, Sauer U, Zamboni N. Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism. Analytical Chemistry 2009; 81:2135-2143.

29. Buschlen S, Amillet JM, Guiard B, Fournier A, Marcireau C, Bolotin-Fukuhara M. The S-cerevisiae HAP complex, a key regulator of mitochondrial function, coordinates nuclear and mitochondrial gene expression. Comparative and Functional Genomics 2003; 4:37-46.

30. Buziol S, Bashir I, Baumeister A, Claassen W, Noisommit-Rizzi N, Mailinger W, Reuss M. New bioreactor-coupled rapid stopped-flow sampling technique for measurements of metabolite dynamics on a subsecond time scale. Biotechnology and Bioengineering 2002; 80:632-636.

31. Canelas AB, Ras C, ten Pierick A, Van Dam JC, Heijnen JJ, Van Gulik WM. Leakage-free rapid quenching technique for yeast metabolomics. Metabolomics 2008; 4:226-239.

32. Canelas AB, ten Pierick A, Ras C, Seifar RM, Van Dam JC, Van Gulik WM, Heijnen JJ. Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics. Analytical Chemistry 2009; 81:7379-7389.

33. Canelas AB, Van Gulik WM, Heijnen JJ. Determination of the cytosolic free NAD/NADH ratio in Saccharomyces cerevisiae under steady-state and highly dynamic conditions. Biotechnology and Bioengineering 2008; 100:734-743.

34. Canonaco F, Schlattner U, Pruett PS, Wallimann T, Sauer U. Functional expression of phosphagen kinase systems confers resistance to transient stresses in Saccharomyces cerevisiae by buffering the ATP pool. Journal of Biological Chemistry 2002; 277:31303-31309.

35. Castrillo JI, Hayes A, Mohammed S, Gaskell SJ, Oliver SG. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. Phytochemistry 2003; 62:929-937.

36. Chaturvedi V, Bartiss A, Wong B. Expression of bacterial mtlD in Saccharomyces cerevisiae results in mannitol synthesis and protects a glycerol-defective mutant from high-salt and oxidative stress. Journal of Bacteriology 1997; 179:157-162.

37. Cipollina C, ten PA, Canelas AB, Seifar RM, van Maris AJ, Van Dam JC, Heijnen JJ. A comprehensive method for the quantification of the non-oxidative pentose phosphate pathway intermediates in Saccharomyces cerevisiae by GC-IDMS. J Chromatogr B Analyt Technol Biomed Life Sci 2009; 877:3231-3236.

38. Cipollina C, van den Brink J, ran-Lapujade P, Pronk JT, Vai M, de Winde JH. Revisiting the role of yeast Sfp1 in ribosome biogenesis and cell size control: a chemostat study. Microbiology-Sgm 2008; 154:337-346.

39. Cleland WW. Kinetics of Enzyme-Catalyzed Reactions with 2 Or More Substrates Or Products .1. Nomenclature and Rate Equations. Biochimica et Biophysica Acta 1963; 67:104-&.

40. Cole HA, Wimpenny JW, Hughes DE. Atp Pool in Escherichia Coli .I. Measurement of Pool Using A Modified Luciferase Assay. Biochimica et Biophysica Acta 1967; 143:445-&.

41. Costenoble R, Adler L, Niklasson C, Liden G. Engineering of the metabolism of Saccharomyces cerevisiae for anaerobic production of mannitol. FEMS Yeast Res 2003; 3:17-25.

42. Daran-Lapujade P, Jansen MLA, Daran JM, van Gulik W, de Winde JH, Pronk JT. Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae - A chemostat culture study. Journal of Biological Chemistry 2004; 279:9125-9138.

43. Daran-Lapujade P, Rossell S, Van Gulik WM, Luttik MAH, de Groot MJL, Slijper M, Heck AJR, Daran JM, de Winde JH, Westerhoff HV, Pronk JT, Bakker BM. The fluxes through glycolytic enzymes in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels. Proceedings of the National Academy of Sciences of the United States of America 2007; 104:15753-15758.

44. David R.Lide. Concentrative Properties of Aqueous Solutions: Density, Refractive Index, Freezing Point Depression, and Viscosity. In: David R.Lide, editor. *CRC Handbook of Chemistry and Physics, Internet Version 2007, (87th Edition).* Boca Raton, FL: Taylor and Francis; 2007.

45. de Koning W, van Dam K. A Method for the Determination of Changes of Glycolytic Metabolites in Yeast on A Subsecond Time Scale Using Extraction at Neutral Ph. Analytical Biochemistry 1992; 204:118-123.

46. Deuschle K, Okumoto S, Fehr M, Looger LL, Kozhukh L, Frommer WB. Construction and optimization of a family of genetically encoded metabolite sensors by semirational protein engineering. Protein Science 2005; 14:2304-2314.

47. Diderich JA, Schepper M, van Hoek P, Luttik MAH, van Dijken JP, Pronk JT, Klaassen P, Boelens HFM, de Mattos RJT, van Dam K, Kruckeberg AL. Glucose uptake kinetics and transcription of HXT genes chemostat cultures of Saccharomyces cerevisiae. Journal of Biological Chemistry 1999; 274:15350-15359.

48. Droge W. Free radicals in the physiological control of cell function. Physiological Reviews 2002; 82:47-95.

49. Entian KD, Kotter P. Yeast mutant and plasmid collections. Yeast Gene Analysis 1998; 26:431-449.

50. Ertugay N, Hamamci H. Continuous cultivation of bakers' yeast: Change in cell composition at different dilution rates and effect of heat stress on trehalose level. Folia Microbiologica 1997; 42:463-467.

51. Ewald JC, Heux S, Zamboni N. High-throughput quantitative metabolomics: workflow for cultivation, quenching, and analysis of yeast in a multiwell format. Anal Chem 2009; 81:3623-3629.

52. Faijes M, Mars AE, Smid EJ. Comparison of quenching and extraction methodologies for metabolome analysis of Lactobacillus plantarum. Microbial Cell Factories 2007; 6.

53. Farre EM, Tiessen A, Roessner U, Geigenberger P, Trethewey RN, Willmitzer L. Analysis of the compartmentation of glycolytic intermediates, nucleotides, sugars, organic acids, amino acids, and sugar alcohols in potato tubers using a nonaqueous fractionation method. Plant Physiology 2001; 127:685-700.

54. Faupel RP, Seitz HJ, Tarnowsk W, Thiemann V, Weiss C. Problem of Tissue Sampling from Experimental-Animals with Respect to Freezing Technique, Anoxia, Stress and Narcosis - New Method for Sampling Rat-Liver Tissue and Physiological Values of Glycolytic Intermediates and Related Compounds. Archives of Biochemistry and Biophysics 1972; 148:509-&.

55. Fleming RMT, Thiele I, Nasheuer HP. Quantitative assignment of reaction directionality in constraint-based models of metabolism: Application to Escherichia coli. Biophysical Chemistry 2009; 145:47-56.

56. Folch J, Lees M, Stanley GHS. A Simple Method for the Isolation and Purification of Total Lipides from Animal Tissues. Journal of Biological Chemistry 1957; 226:497-509.

57. Francois J, Neves MJ, Hers HG. The Control of Trehalose Biosynthesis in Saccharomyces-Cerevisiae - Evidence for A Catabolite Inactivation and Repression of Trehalose-6-Phosphate Synthase and Trehalose-6-Phosphate Phosphatase. Yeast 1991; 7:575-587.

58. Fuerst R, Wagner RP. An Analysis of the Free Intracellular Amino Acids of Certain Strains of Neurospora. Archives of Biochemistry and Biophysics 1957; 70:311-326.

59. Gale EF. The Assimilation of Amino-Acids by Bacteria .1. the Passage of Certain Amino-Acids Across the Cell Wall and Their Concentration in the Internal Environment of Streptococcus-Faecalis. Journal of General Microbiology 1947; 1:53-76.

60. Gao H, Leary JA. Multiplex inhibitor screening and kinetic constant determinations for yeast hexokinase using mass spectrometry based assays. Journal of the American Society for Mass Spectrometry 2003; 14:173-181.

61. Ge X, Sirich TL, Beyer MK, Desaire H, Leary JA. A strategy for the determination of enzyme kinetics using electrospray ionization with an ion trap mass spectrometer. Analytical Chemistry 2001; 73:5078-5082.

62. Gerber SA, Rush J, Stemman O, Kirschner MW, Gygi SP. 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 2003; 100:6940-6945.

63. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, Arkin AP, Astromoff A, El Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss

M, Davis K, Deutschbauer A, Entian KD, Flaherty P, Foury F, Garfinkel DJ, Gerstein M, Gotte D, Guldener U, Hegemann JH, Hempel S, Herman Z, Jaramillo DF, Kelly DE, Kelly SL, Kotter P, LaBonte D, Lamb DC, Lan N, Liang H, Liao H, Liu L, Luo CY, Lussier M, Mao R, Menard P, Ooi SL, Revuelta JL, Roberts CJ, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker DD, Sookhai-Mahadeo S, Storms RK, Strathern JN, Valle G, Voet M, Volckaert G, Wang CY, Ward TR, Wilhelmy J, Winzeler EA, Yang YH, Yen G, Youngman E, Yu KX, Bussey H, Boeke JD, Snyder M, Philippsen P, Davis RW, Johnston M. Functional profiling of the Saccharomyces cerevisiae genome. Nature 2002; 418:387-391.

64. Gietz RD, Woods RA. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Guide to Yeast Genetics and Molecular and Cell Biology, Pt B 2002; 350:87-96.

65. Gillies RJ, Ugurbil K, Denhollander JA, Shulman RG. P-31 Nmr-Studies of Intracellular Ph and Phosphate-Metabolism During Cell-Division Cycle of Saccharomyces-Cerevisiae. Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences 1981; 78:2125-2129.

66. Goldberg RN, Tewari YB, Bhat TN. Thermodynamics of enzyme-catalyzed reactions - a database for quantitative biochemistry. Bioinformatics 2004; 20:2874-2877.

67. Gonzalez B, Francois J, Renaud M. A rapid and reliable method for metabolite extraction in yeast using boiling buffered ethanol. Yeast 1997; 13:1347-1355.

68. Gunawardena J. Models in Systems Biology: The Parameter Problem and the Meanings of Robustness. In: Lodhi HM, Muggleton SH, editors. Elements of Computational Systems Biology. Wiley; 2010.

69. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. Plos Computational Biology 2007; 3:1871-1878.

70. Hajjaj H, Blanc PJ, Goma G, Francois J. Sampling techniques and comparative extraction procedures for quantitative determination of intra- and extracellular metabolites in filamentous fungi. Fems Microbiology Letters 1998; 164:195-200.

71. Hancock R. The Intracellular Amino Acids of Staphylococcus-Aureus - Release and Analysis. Biochimica et Biophysica Acta 1958; 28:402-412.

72. Hans MA, Heinzle E, Wittmann C. Quantification of intracellular amino acids in batch cultures of Saccharomyces cerevisiae. Applied Microbiology and Biotechnology 2001; 56:776-779.

73. Harrison DE, Maitra PK. Control of Respiration and Metabolism in Growig Klebsiella Aerogenes - Role of Adenine Nucleotides. Biochemical Journal 1969; 112:647-&.

74. Heijnen JJ. Approximative kinetic formats used in metabolic network modeling. Biotechnology and Bioengineering 2005; 91:534-545.

75. Heinrich R, Schuster S. The regulation of cellular systems. Chapman; 1996.

76. Henry CS, Jankowski MD, Broadbelt LJ, Hatzimanikatis V. Genome-scale thermodynamic analysis of Escherichia coli metabolism. Biophysical Journal 2006; 90:1453-1461.

77. Hiller J, Franco-Lara E, Weuster-Botz D. Metabolic profiling of Escherichia coli cultivations: evaluation of extraction and metabolite analysis procedures. Biotechnology Letters 2007; 29:1169-1178.

78. Hockley SL, Mathijs K, Staal YCM, Brewer D, Giddings I, van Delft JHM, Phillips DH. Interlaboratory and Interplatform Comparison of Microarray Gene Expression Analysis of HepG2 Cells Exposed to Benzo(a)pyrene. Omics-A Journal of Integrative Biology 2009; 13:115-125.

79. Holzer H, Schultz G, Lynen F. Bestimmung des Quotienten Dpnh-Dpn in Lebenden Hefezellen Durch Analyse Stationarer Alkohol-Konzentrationen und Acetaldehyd-Konzentrationen. Biochemische Zeitschrift 1956; 328:252-263.

80. Hommes FA. Oscillatory Reductions of Pyridine Nucleotides During Anaerobic Glycolysis in Brewers Yeast. Archives of Biochemistry and Biophysics 1964; 108:36-46.

81. Hoskisson PA, Hobbs G. Continuous culture - making a comeback? Microbiology-Sgm 2005; 151:3153-3159.

82. Hynne R, Dano S, Sorensen PG. Full-scale model of glycolysis in Saccharomyces cerevisiae. Biophysical Chemistry 2001; 94:121-163.

83. Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, Eng JK, Bumgarner R, Goodlett DR, Aebersold R, Hood L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science 2001; 292:929-934.

84. Jamshidi N, Palsson BO. Formulating genome-scale kinetic models in the post-genome era. Molecular Systems Biology 2008; 4.

85. Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V. Group contribution method for thermodynamic analysis of complex metabolic networks. Biophysical Journal 2008; 95:1487-1499.

86. Jansen MLA, Diderich JA, Mashego M, Hassane A, de Winde JH, ran-Lapujade P, Pronk JT. Prolonged selection in aerobic, glucose-limited chemostat cultures of Saccharomyces cerevisiae causes a partial loss of glycolytic capacity. Microbiology-Sgm 2005; 151:1657-1669.

87. Jeneson JAL, Westerhoff HV, Brown TR, Vanechteld CJA, Berger R. Quasi-Linear Relationship Between Gibbs Free-Energy of Atp Hydrolysis and Power Output in Human Forearm Muscle. American Journal of Physiology-Cell Physiology 1995; 37:C1474-C1484.

88. Jennings P, Aydin S, Bennett J, McBride R, Weiland C, Tuite N, Gruber LN, Perco P, Gaora PO, Ellinger-Ziegelbauer H, Ahr HJ, Van Kooten C, Doha MR, Prieto P, Ryan MP, Pfaller W, McMorrow T. Inter-laboratory comparison of human renal proximal tubule (HK-2) transcriptome alterations due to Cyclosporine A exposure and medium exhaustion. Toxicology in Vitro 2009; 23:486-499.

89. Jernejc K. Comparison of different methods for metabolite extraction from Aspergillus niger mycelium. Acta Chimica Slovenica 2004; 51:567-578.

90. Kao HP, Abney JR, Verkman AS. Determinants of the Translational Mobility of A Small Solute in Cell Cytoplasm. Journal of Cell Biology 1993; 120:175-184.

91. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahren D, Tsoka S, Darzentas N, Kunin V, Lopez-Bigas N. Expansion of the BioCyc collection of pathway/genome databases to 160 genomes. Nucleic Acids Research 2005; 33:6083-6089.

92. Kiser RC, Niehaus WG. Purification and Kinetic Characterization of Mannitol-1-Phosphate Dehydrogenase from Aspergillus-Niger. Archives of Biochemistry and Biophysics 1981; 211:613-621.

93. Kitano H. Systems biology: A brief overview. Science 2002; 295:1662-1664.

94. Krebs HA, Wiggins D, Stubbs M, Sols A, Bedoya F. Studies on the Mechanism of the Anti-Fungal Action of Benzoate. Biochemical Journal 1983; 214:657-663.

95. Kresnowati MTAP, Suarez-Mendez C, Groothuizen MK, Van Winden WA, Heijnen JJ. Measurement of fast dynamic intracellular pH in Saccharomyces cerevisiae, using benzoic acid pulse. Biotechnology and Bioengineering 2007; 97:86-98.

96. Kresnowati MTAP, Van Winden WA, Almering MJH, ten Pierick A, Ras C, Knijnenburg TA, ran-Lapujade P, Pronk JT, Heijnen JJ, Daran JM. When transcriptome meets metabolome: fast cellular responses of yeast to sudden relief of glucose limitation. Molecular Systems Biology 2006.

97. Kresnowati MTAP, Van Winden WA, Heijnen JJ. Determination of elasticities, concentration and flux control coefficients from transient metabolite data using linlog kinetics. Metabolic Engineering 2005; 7:142-153.

98. Kuenzi MT, Fiechter A. Regulation of Carbohydrate Composition of Saccharomyces-Cerevisiae Under Growth Limitation. Archiv fur Mikrobiologie 1972; 84:254-&.

99. Kummel A, Panke S, Heinemann M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. Molecular Systems Biology 2006; 2.

100. Lange HC, Eman M, van Zuijlen G, Visser D, Van Dam JC, Frank J, de Mattos MJT, Heijnen JJ. Improved rapid sampling for in vivo kinetics of intracellular metabolites in Saccharomyces cerevisiae. Biotechnology and Bioengineering 2001; 75:406-415.

101. Lange HC, Heijnen JJ. Statistical reconciliation of the elemental and molecular biomass composition of Saccharomyces cerevisiae. Biotechnology and Bioengineering 2001; 75:334-344.

102. Larsson CM, Olsson T. Firefly Assay of Adenine-Nucleotides from Algae - Comparison of Extraction Methods. Plant and Cell Physiology 1979; 20:145-155.

103. Le Novere N, Finney A, Hucka M, Bhalla US, Campagne F, Collado-Vides J, Crampin EJ, Halstead M, Klipp E, Mendes P, Nielsen P, Sauro H, Shapiro B, Snoep JL, Spence HD, Wanner BL. Minimum information requested in the annotation of biochemical models (MIRIAM). Nature Biotechnology 2005; 23:1509-1515.

104. Lin SJ, Ford E, Haigis M, Liszt G, Guarente L. Calorie restriction extends yeast life span by lowering the level of NADH. Genes & Development 2004; 18:12-16.

105. Link H, Anselment B, Weuster-Botz D. Leakage of adenylates during cold methanol/glycerol quenching of Escherichia coli. Metabolomics 2008; 4:240-247.

106. Loret MO, Pedersen L, Francois J. Revised procedures for yeast metabolites extraction: application to a glucose pulse to carbon-limited yeast cultures, which reveals a transient activation of the purine salvage pathway. Yeast 2007; 24:47-60.

107. Lundin A, Thore A. Comparison of Methods for Extraction of Bacterial Adenine-Nucleotides Determined by Firefly Assay. Applied Microbiology 1975; 30:713-721.

108. Luttik MAH, Vuralhan Z, Suir E, Braus GH, Pronk JT, Daran JM. Alleviation of feedback inhibition in Saccharomyces cerevisiae aromatic amino acid biosynthesis: Quantification of metabolic impact. Metabolic Engineering 2008; 10:141-153.

109. Maharjan RP, Ferenci T. Global metabolite analysis: the influence of extraction methodology on metabolome profiles of Escherichia coli. Analytical Biochemistry 2003; 313:145-154.

110. 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.

111. Martinezforce E, Benitez T. Effects of Varying Media, Temperature, and Growth-Rates on the Intracellular Concentrations of Yeast Amino-Acids. Biotechnology Progress 1995; 11:386-392.

112. Mashego MR, Gulik WM, Heijnen JJ. Metabolome dynamic responses of Saccharomyces cerevisiae to simultaneous rapid perturbations in external electron acceptor and electron donor. Fems Yeast Research 2007; 7:48-66.

113. Mashego MR, Jansen MLA, Vinke JL, Van Gulik WM, Heijnen JJ. Changes in the metabolome of Saccharomyces cerevisiae associated with evolution in aerobic glucose-limited chemostats. Fems Yeast Research 2005; 5:419-430.

114. Mashego MR, Rumbold K, De Mey M, Vandamme E, Soetaert W, Heijnen JJ. Microbial metabolomics: past, present and future methodologies. Biotechnology Letters 2007; 29:1-16.

115. Mashego MR, Van Gulik WM, Vinke JL, Heijnen JJ. Critical evaluation of sampling techniques for residual glucose determination in carbon-limited chemostat culture of Saccharomyces cerevisiae. Biotechnology and Bioengineering 2003; 83:395-399.

116. Mashego MR, Van Gulik WM, Vinke JL, Visser D, Heijnen JJ. In vivo kinetics with rapid perturbation experiments in Saccharomyces cerevisiae using a second-generation BioScope. Metabolic Engineering 2006; 8:370-383.

117. Mashego MR, Wu L, Van Dam JC, Ras C, Vinke JL, Van Winden WA, Van Gulik WM, Heijnen JJ. MIRACLE: mass isotopomer ratio analysis of U-C-13-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. Biotechnology and Bioengineering 2004; 85:620-628.

118. Maskow T, von Stockar U. How reliable are thermodynamic feasibility statements of biochemical pathways? Biotechnology and Bioengineering 2005; 92:223-230.

119. Minton AP. How can biochemical reactions within cells differ from those in test tubes? Journal of Cell Science 2006; 119:2863-2869.

120. More N, Daniel RM, Petach HH. The Effect of Low-Temperatures on Enzyme-Activity. Biochemical Journal 1995; 305:17-20.

121. Moritz B, Striegel K, de Graaf AA, Sahm H. Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from Corynebacterium glutamicum and their application for predicting pentose phosphate pathway flux in vivo. European Journal of Biochemistry 2000; 267:3442-3452.

122. Morozov VN, Morozova TY, Johnson KL, Naylor S. Parallel determination of multiple protein metabolite interactions using cell extract, protein microarrays and mass spectrometric detection. Rapid Communications in Mass Spectrometry 2003; 17:2430-2438.

123. Mulet JM, Alejandro S, Romero C, Serrano R. The trehalose pathway and intracellular glucose phosphates as modulators of potassium transport and general cation homeostasis in yeast. Yeast 2004; 21:569-582.

124. Nikerel IE, Canelas AB, Jol S., Verheijen PJT, Heijnen JJ. Construction of kinetic models for metabolic reaction networks: lessons learned in analyzing short term stimulus response data. MATHMOD Conference Proocedings 2009;805-814.

125. Nikerel IE, Van Winden WA, Van Gulik WM, Heijnen JJ. A method for estimation of elasticities in metabolic networks using steady state and dynamic metabolomics data and linlog kinetics. Bmc Bioinformatics 2006; 7.

126. Nikerel IE, Van Winden WA, Verheijen PJT, Heijnen JJ. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. Metabolic Engineering 2009; 11:20-30.

127. Novotny MJ, Reizer J, Esch F, Saier MH. Purification and Properties of D-Mannitol-1-Phosphate Dehydrogenase and D-Glucitol-6-Phosphate Dehydrogenase from Escherichia-Coli. Journal of Bacteriology 1984; 159:986-990.

128. Oldiges M, Takors R. Applying metabolic profiling techniques for stimulus-response experiments: Chances and pitfalls. Technology Transfer in Biotechnology: from Lab to Industry to Production 2005; 92:173-196.

129. Oliveira AP, Patil KR, Nielsen J. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. Bmc Systems Biology 2008; 2.

130. Oliver SG, Winson MK, Kell DB, Baganz F. Systematic functional analysis of the yeast genome. Trends in Biotechnology 1998; 16:373-378.

131. Onodera J, Ohsumi Y. Autophagy is required for maintenance of amino acid levels and protein synthesis under nitrogen starvation. Journal of Biological Chemistry 2005; 280:31582-31586.

132. Onsager L. Reciprocal relations in irreversible processes. I. Physical Review 1931; 37:405-426.

133. Onsager L. Reciprocal relations in irreversible processes. II. Physical Review 1931; 38:2265-2279.

134. Orij R, Postmus J, Ter Beek A, Brul S, Smits GJ. In vivo measurement of cytosolic and mitochondrial pH using a pH-sensitive GFP derivative in Saccharomyces cerevisiae reveals a relation between intracellular pH and growth. Microbiology-Sgm 2009; 155:268-278.

135. Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. Proceedings of the National Academy of Sciences of the United States of America 2005; 102:2685-2689.

136. Picotti P, Bodenmiller B, Mueller LN, Domon B, Aebersold R. Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. Cell 2009; 138:795-806.

137. Piper MDW, ran-Lapujade P, Bro C, Regenberg B, Knudsen S, Nielsen J, Pronk JT. Reproducibility of oligonucleotide microarray transcriptome analyses - An interlaboratory comparison using chemostat cultures of Saccharomyces cerevisiae. Journal of Biological Chemistry 2002; 277:37001-37008.

138. Postma E, Verduyn C, Scheffers WA, Vandijken JP. Enzymic Analysis of the Crabtree Effect in Glucose-Limited Chemostat Cultures of Saccharomyces-Cerevisiae. Applied and Environmental Microbiology 1989; 55:468-477.

139. Postmus J, Canelas AB, Bouwman J, Bakker BM, van Gulik W, de Mattos MJT, Brul S, Smits GJ. Quantitative analysis of the high temperature-induced glycolytic flux increase in Saccharomyces cerevisiae reveals dominant metabolic regulation. Journal of Biological Chemistry 2008; 283:23524-23532.

140. Pratt JM, Simpson DM, Doherty MK, Rivers J, Gaskell SJ, Beynon RJ. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. Nature Protocols 2006; 1:1029-1043.

141. Rabinowitz JD, Kimball E. Acidic acetonitrile for cellular metabolome extraction from Escherichia coli. Analytical Chemistry 2007; 79:6167-6173.

142. Ramachandran N, Larson DN, Stark PRH, Hainsworth E, LaBaer J. Emerging tools for real-time label-free detection of interactions on functional protein microarrays. Febs Journal 2005; 272:5412-5425.

143. Rautio JJ, Kataja K, Satokani R, Penttila M, Soderlund H, Saloheimo M. Rapid and multiplexed transcript analysis of microbial cultures using capillary electophoresis-detectable oligonucleotide probe pools. Journal of Microbiological Methods 2006; 65:404-416.

144. Regenberg B, Grotkjaer T, Winther O, Fausboll A, Akesson M, Bro C, Hansen LK, Brunak S, Nielsen J. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biology 2006; 7.

145. Regev-Rudzki N, Battat E, Goldberg I, Pines O. Dual localization of fumarase is dependent on the integrity of the glyoxylate shunt. Molecular Microbiology 2009; 72:297-306.

146. Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hon GC, Myers CL, Parsons A, Friesen H, Oughtred R, Tong A, Stark C, Ho Y, Botstein D, Andrews B, Boone C, Troyanskya OG, Ideker T, Dolinski K, Batada NN, Tyers M. Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. J Biol 2006; 5:11.

147. Richard P, Teusink B, Hemker MB, Vandam K, Westerhoff HV. Sustained oscillations in free-energy state and hexose phosphates in yeast. Yeast 1996; 12:731-740.

148. Riens B, Lohaus G, Heineke D, Heldt HW. Amino-Acid and Sucrose Content Determined in the Cytosolic, Chloroplastic, and Vacuolar Compartments and in the Phloem Sap of Spinach Leaves. Plant Physiology 1991; 97:227-233.

149. Rizzi M, Baltes M, Theobald U, Reuss M. In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae .2. Mathematical model. Biotechnology and Bioengineering 1997; 55:592-608.

150. Rose IA. State of Magnesium in Cells As Estimated from Adenylate Kinase Equilibrium. Proceedings of the National Academy of Sciences of the United States of America 1968; 61:1079-&.

151. Rottenberg H. Thermodynamic Description of Enzyme-Catalyzed Reactions - Linear Relation Between Reaction-Rate and Affinity. Biophysical Journal 1973; 13:503-511.

152. Ruijter GJG, Visser J. Determination of intermediary metabolites in Aspergillus niger. Journal of Microbiological Methods 1996; 25:295-302.

153. Saez MJ, Lagunas R. Determination of Intermediary Metabolites in Yeast - Critical-Examination of Effect of Sampling Conditions and Recommendations for Obtaining True Levels. Molecular and Cellular Biochemistry 1976; 13:73-78.

154. Saito N, Robert M, Kitamura S, Baran R, Soga T, Mori H, Nishioka T, Tomita M. Metabolomics approach for enzyme discovery. Journal of Proteome Research 2006; 5:1979-1987.

155. Saks V, Dos Santos P, Gellerich FN, Diolez P. Quantitative studies of enzyme-substrate compartmentation, functional coupling and metabolic channelling in muscle cells. Molecular and Cellular Biochemistry 1998; 184:291-307.

156. Schaaff I, Heinisch J, Zimmermann FK. Overproduction of Glycolytic-Enzymes in Yeast. Yeast 1989; 5:285-290.

157. Schaub J, Schiesling C, Reuss M, Dauner M. Integrated sampling procedure for metabolome analysis. Biotechnology Progress 2006; 22:1434-1442.

158. Seifar RM, Ras C, van Dam JC, van Gulik WM, Heijnen JJ, van Winden WA. Simultaneous quantification of free nucleotides in complex biological samples using ion pair reversed phase liquid chromatography isotope dilution tandem mass spectrometry. Analytical Biochemistry 2009; 388:213-219.

159. Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, Cazares LH, Chan DW, Grizzle WE, Izbicka E, Kagan J, Malik G, McLerran D, Moul JW, Partin A, Prasanna P, Rosenzweig J, Sokoll LJ, Srivastava S, Srivastava S, Thompson I, Welsh MJ, White N, Winget M, Yasui Y, Zhang Z, Zhu L. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. Clinical Chemistry 2005; 51:102-112.

160. Shearer G, Lee JC, Koo J, Kohl DH. Quantitative estimation of channeling from early glycolytic intermediates to CO2 in intact Escherichia coli. Febs Journal 2005; 272:3260-3269.

161. Sillje HHW, Paalman JWG, ter Schure EG, Olsthoorn SQB, Verkleij AJ, Boonstra J, Verrips CT. Function of trehalose and glycogen in cell cycle progression and cell viability in Saccharomyces cerevisiae. Journal of Bacteriology 1999; 181:396-400.

162. Smits HP, Cohen A, Buttler T, Nielsen J, Olsson L. Cleanup and analysis of sugar phosphates in biological extracts by using solid-phase extraction and anion-exchange chromatography with pulsed amperometric detection. Analytical Biochemistry 1998; 261:36-42.

163. Smits HP, Hauf J, Muller S, Hobley TJ, Zimmermann FK, Hahn-Hagerdal B, Nielsen J, Olsson L. Simultaneous overexpression of enzymes of the lower part of glycolysis can enhance the fermentative capacity of Saccharomyces cerevisiae. Yeast 2000; 16:1325-1334.

164. Stubbs M, Veech RL, Krebs HA. Control of Redox State of Nicotinamide-Adenine Dinucleotide Couple in Rat-Liver Cytoplasm. Biochemical Journal 1972; 126:59-65.

165. Taylor CF, Paton NW, Lilley KS, Binz PA, Julian RK, Jones AR, Zhu WM, Apweiler R, Aebersold R, Deutsch EW, Dunn MJ, Heck AJR, Leitner A, Macht M, Mann M, Martens L, Neubert TA, Patterson SD, Ping PP, Seymour SL, Souda P, Tsugita A, Vandekerckhove J, Vondriska TM, Whitelegge JP, Wilkins MR, Xenarios I, Yates JR, Hermjakob H. The minimum information about a proteomics experiment (MIAPE). Nature Biotechnology 2007; 25:887-893.

166. Taymaz-Nikerel H, De Mey M, Ras C, ten Pierick A, Seifar RM, Van Dam JC, Heijnen JJ, Van Gulik WM. Development and application of a differential method for reliable metabolome analysis in Escherichia coli. Analytical Biochemistry 2009; 386:9-19.

167. Teleman A, Richard P, Toivari M, Penttilla M. Identification and quantitation of phosphorus metabolites in yeast neutral pH extracts by nuclear magnetic resonance spectroscopy. Analytical Biochemistry 1999; 272:71-79.

168. Teschner W, Serre MC, Garel JR. Enzymatic-Properties, Renaturation and Metabolic Role of Mannitol-1-Phosphate Dehydrogenase from Escherichia-Coli. Biochimie 1990; 72:33-40.

169. Teusink B, Passarge J, Reijenga CA, Esgalhado E, van der Weijden CC, Schepper M, Walsh MC, Bakker BM, van Dam K, Westerhoff HV, Snoep JL. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. European Journal of Biochemistry 2000; 267:5313-5329.

170. Theobald U, Mailinger W, Baltes M, Rizzi M, Reuss M. In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae .1. Experimental observations. Biotechnology and Bioengineering 1997; 55:305-316.

171. Theobald U, Mailinger W, Reuss M, Rizzi M. In-Vivo Analysis of Glucose-Induced Fast Changes in Yeast Adenine-Nucleotide Pool Applying A Rapid Sampling Technique. Analytical Biochemistry 1993; 214:31-37.

172. Van Dam JC, Eman MR, Frank J, Lange HC, van Dedem GWK, Heijnen SJ. Analysis of glycolytic intermediates in Saccharomyces cerevisiae using anion exchange chromatography and electrospray ionization with tandem mass spectrometric detection. Analytica Chimica Acta 2002; 460:209-218.

173. van den Brink J, Canelas AB, Van Gulik WM, Pronk JT, Heijnen JJ, de Winde JH, ran-Lapujade P. Dynamics of glycolytic regulation during adaptation of Saccharomyces cerevisiae to fermentative metabolism. Applied and Environmental Microbiology 2008; 74:5710-5723.

174. van der Meer R, Westerhoff HV, van Dam K. Linear Relation Between Rate and Thermodynamic Force in Enzyme-Catalyzed Reactions. Biochimica et Biophysica Acta 1980; 591:488-493.

175. van der Werf MJ, Takors R, Smedsgaard J, Nielsen J, Ferenci T, Portais JC, Wittmann C, Hooks M, Tomassini A, Oldiges M, Fostel J, Sauer U. Standard reporting requirements for biological samples in metabolomics experiments: microbial and in vitro biology experiments. Metabolomics 2007; 3:189-194.

176. van Dijken JP, Bauer J, Brambilla L, Duboc P, Francois JM, Gancedo C, Giuseppin MLF, Heijnen JJ, Hoare M, Lange HC, Madden EA, Niederberger P, Nielsen J, Parrou JL, Petit T, Porro D, Reuss M, van Riel N, Rizzi M, Steensma HY, Verrips CT, Vindelov J, Pronk JT. An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains. Enzyme and Microbial Technology 2000; 26:706-714.

177. van Eunen K, Bouwman J, Lindenbergh A, Westerhoff HV, Bakker BM. Time-dependent regulation analysis dissects shifts between metabolic and gene-expression regulation during nitrogen starvation in baker's yeast. Febs Journal 2009; 276:5521-5536.

178. van Hoek P, van Dijken JP, Pronk JT. Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of Saccharomyces cerevisiae. Enzyme and Microbial Technology 2000; 26:724-736.

179. vanGulik WM, Heijnen JJ. A Metabolic Network Stoichiometry Analysis of Microbial-Growth and Product Formation. Biotechnology and Bioengineering 1995; 48:681-698.

180. Veech RL. A humble hexose monophosphate pathway metabolite regulates short- and long-term control of lipogenesis. Proceedings of the National Academy of Sciences of the United States of America 2003; 100:5578-5580.

181. Veech RL. The determination of the redox states and phosphorylation potential in living tissues and their relationship to metabolic control of disease phenotypes. Biochemistry and Molecular Biology Education 2006; 34:168-179.

182. Veech RL, Egglesto LV, Krebs HA. Redox State of Free Nicotinamide-Adenine Dinucleotide Phosphate in Cytoplasm of Rat Liver. Biochemical Journal 1969; 115:609-619.

183. Veech RL, Raijman L, Krebs HA. Equilibrium Relations Between Cytoplasmic Adenine Nucleotide System and Nicotinamide-Adenine Nucleotide System in Rat Liver. Biochemical Journal 1970; 117:499-503.

184. Verduyn C. Energetic aspects of metabolic fluxes in yeasts. Dissertation, TU Delft, The Netherlands; 1992.

185. Verduyn C, Postma E, Scheffers WA, Vandijken JP. Effect of Benzoic-Acid on Metabolic Fluxes in Yeasts - A Continuous-Culture Study on the Regulation of Respiration and Alcoholic Fermentation. Yeast 1992; 8:501-517.

186. Verkman AS. Solute and macromolecule diffusion in cellular aqueous compartments. Trends in Biochemical Sciences 2002; 27:27-33.

187. Villas-Boas SG, Bruheim P. Cold glycerol-saline: The promising quenching solution for accurate intracellular metabolite analysis of microbial cells. Analytical Biochemistry 2007; 370:87-97.

188. Villas-Boas SG, Hojer-Pedersen J, Akesson M, Smedsgaard J, Nielsen J. Global metabolite analysis of yeast: evaluation of sample preparation methods. Yeast 2005; 22:1155-1169.

189. Visser D, van Zuylen GA, Van Dam JC, Eman MR, Proll A, Ras C, Wu L, Van Gulik WM, Heijnen JJ. Analysis of in vivo kinetics of glycolysis in aerobic Saccharomyces cerevisiae by application of glucose and ethanol pulses. Biotechnology and Bioengineering 2004; 88:157-167.

190. Visser D, van Zuylen GA, Van Dam JC, Oudshoorn A, Eman MR, Ras C, Van Gulik WM, Frank J, van Dedem GWK, Heijnen JJ. Rapid sampling for analysis of in vivo kinetics using the BioScope: A system for continuous-pulse experiments. Biotechnology and Bioengineering 2002; 79:674-681.

191. von Stockar U, Maskow T, Liu JS, Marison IW, Patino R. Thermodynamics of microbial growth and metabolism: An analysis of the current situation. Journal of Biotechnology 2006; 121:517-533.

192. Wahl SA, Noh K, Wiechert W. C-13 labeling experiments at metabolic nonstationary conditions: An exploratory study. Bmc Bioinformatics 2008; 9.

193. Wang QZ, Yang YD, Chen X, Zhao XM. Comparisons of different extraction methods in Escherichia coli metabolome analysis. Chinese Journal of Analytical Chemistry 2006; 34:1295-1298.

194. Waterfall JJ, Casey FP, Gutenkunst RN, Brown KS, Myers CR, Brouwer PW, Elser V, Sethna JP. Sloppy-model universality class and the Vandermonde matrix. Physical Review Letters 2006; 97.

195. Weibel KE, Mor JR, Fiechter A. Rapid Sampling of Yeast-Cells and Automated Assays of Adenylate, Citrate, Pyruvate and Glucose-6-Phosphate Pools. Analytical Biochemistry 1974; 58:208-216.

196. Westerhoff HV, van Dam K. Thermodynamics and control of biological free-energy transduction. Elsevier; 1987.

197. Wiebe MG, Rintala E, Tamminen A, Simolin H, Salusjarvi L, Toivari M, Kokkonen JT, Kiuru J, Ketola RA, Jouhten P, Huuskonen A, Maaheimo H, Ruohonen L, Penttila M. Central carbon metabolism of Saccharomyces cerevisiae in anaerobic, oxygen-limited and fully aerobic steady-state conditions and following a shift to anaerobic conditions. Fems Yeast Research 2008; 8:140-154.

198. Williams BJ, Cameron CJ, Workman R, Broeckling CD, Sumner LW, Smith JT. Amino acid profiling in plant cell cultures: An inter-laboratory comparison of CE-MS and GC-MS. Electrophoresis 2007; 28:1371-1379.

199. Williams DH, Lund P, Krebs HA. Redox State of Free Nicotinamide-Adenine Dinucleotide in Cytoplasm and Mitochondria of Rat Liver. Biochemical Journal 1967; 103:514-&.

200. Williamson JR, Chang K, Frangos M, Hasan KS, Ido Y, Kawamura T, Nyengaard JR, Vandenenden M, Kilo C, Tilton RG. Hyperglycemic Pseudohypoxia and Diabetic Complications. Diabetes 1993; 42:801-813.

201. Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis R, Stephens GM, Goodacre R. Global metabolic profiling of Escherichia coli cultures: An evaluation of methods for quenching and extraction of intracellular metabolites. Analytical Chemistry 2008; 80:2939-2948.

202. Winston F, Dollard C, Ricuperohovasse SL. Construction of A Set of Convenient Saccharomyces-Cerevisiae Strains That Are Isogenic to S288C. Yeast 1995; 11:53-55.

203. Wisselink HW, Moers APHA, Mars AE, Hoefnagel MHN, de Vos WM, Hugenholtz J. Overproduction of heterologous mannitol 1-phosphatase: a key factor for engineering mannitol production by Lactococcus lactis. Applied and Environmental Microbiology 2005; 71:1507-1514.

204. Wittmann C, Kromer JO, Kiefer P, Binz T, Heinzle E. Impact of the cold shock phenomenon on quantification of intracellular metabolites in bacteria. Analytical Biochemistry 2004; 327:135-139.

205. Wolff JB, Kaplan NO. D-Mannitol 1-Phosphate Dehydrogenase from Escherichia-Coli. Journal of Biological Chemistry 1956; 218:849-869.

206. Wollenberger A, Ristau O, Schoffa G. Eine Einfache Technik der Extrem Schnellen Abkuhlung Grosserer Gewebestucke. Pflugers Archiv fur Die Gesamte Physiologie des Menschen und der Tiere 1960; 270:399-412.

207. Work E. Chromatographic Investigations of Amino Acids from Micro-Organisms .1. the Amino Acids of Corynebacterium-Diphtheriae. Biochimica et Biophysica Acta 1949; 3:400-411.

208. Wu L, Mashego MR, Proell AM, Vinke JL, Ras C, van Dam J, Van Winden WA, Van Gulik WM, Heijnen JJ. In vivo kinetics of primary metabolism in Saccharomyces cerevisiae studied through prolonged chemostat cultivation. Metabolic Engineering 2006; 8:160-171.

209. Wu L, Mashego MR, Van Dam JC, Proell AM, Vinke JL, Ras C, Van Winden WA, Van Gulik WM, Heijnen JJ. Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly C-13-labeled cell extracts as internal standards. Analytical Biochemistry 2005; 336:164-171.

210. Wu L, van Dam J, Schipper D, Kresnowati MTAP, Proell AM, Ras C, Van Winden WA, Van Gulik WM, Heijnen JJ. Short-term metabolome dynamics and carbon, electron, and ATP balances in chemostat-grown Saccharomyces cerevisiae CEN.PK 113-7D following a glucose pulse. Applied and Environmental Microbiology 2006; 72:3566-3577.

211. Yang WC, Sedlak M, Regnier FE, Mosier N, Ho N, Adamec J. Simultaneous Quantification of Metabolites Involved in Central Carbon and Energy Metabolism Using Reversed-Phase Liquid Chromatography-Mass Spectrometry and in Vitro (13)C Labeling. Anal Chem 2008; 80:9508-9516.

212. Zhang QH, Piston DW, Goodman RH. Regulation of corepressor function by nuclear NADH. Science 2002; 295:1895-1897.

213. Ziegler M. A vital link between energy and signal transduction - Regulatory functions of NAD(P). Febs Journal 2005; 272:4561-4564.

List of publications

Journal articles:

Canelas AB, van Gulik WM, Heijnen JJ (2008) Determination of the cytosolic free NAD/NADH ratio in Saccharomyces cerevisiae under steady-state and highly dynamic conditions, Biotechnology and Bioengineering, 100, 4, 734-743

Canelas AB, Ras C, ten Pierick A, van Dam JC, Heijnen JJ, van Gulik WM (2008) Leakage-free rapid quenching technique for yeast metabolomics, Metabolomics, 4, 3, 226-239

Postmus J, **Canelas AB**, Bouwman J, Bakker BM, van Gulik W, de Mattos MJT, Brul S, Smits GJ (2008) Quantitative analysis of the effect of temperature on glycolytic flux in *Saccharomyces cerevisiae* reveals dominant metabolic regulation, Journal of Biological Chemistry, 283, 35, 23524-23532

van den Brink J, **Canelas AB**, van Gulik WM, et al. (2008) The dynamics of glycolytic regulation during adaptation of Saccharomyces cerevisiae to fermentative metabolism, Applied and Environmental Microbiology, 74, 18, 5710-5723

Canelas AB, ten Pierick A, Ras C, Seifer RM, van Dam JC, van Gulik WM, Heijnen JJ (2009) Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics, Analytical Chemistry, 81, 17, 7379-7389

van Leeuwen M, Buijs NAA, **Canelas AB**, Oudshoorn A, Heijnen JJ, van Gulik WM (2009) The Hagen-Poiseuille pump for parallel fed-batch cultivations in microbioreactors, Chemical Engineering Science, 64, 8, 1877-1884

Cipollina C, ten Pierick A, **Canelas AB**, Seifar RM, van Maris AJA, van Dam JC, Heijnen JJ (2009) A comprehensive method for the quantification of the non-oxidative pentose phosphate pathway intermediates in *Saccharomyces cerevisiae* by GC-IDMS, Journal of Chromatography B, 877, 27, 3231-3236

van Eunen K, Bouwman J, Daran-Lapujade P, Postmus J, **Canelas AB**, Mensonides FIC, Orij R, Tuzun I, van den Brink, [8 others], Westerhoff HV, Bakker BM (2009) Measuring enzyme activities under standardized *in vivo*-like conditions for systems biology, FEBS Journal, 277, 3, 749-760

van Eunen K, Dool P, **Canelas AB**, Kiewiet J, Bouwman J, van Gulik W, Westerhoff HV, Bakker BM (2010) Time-dependent regulation of yeast glycolysis upon nitrogen starvation depends on cell history, IET Systems Biology, 4, 2, 157–168

Canelas AB, Harrison N, Fazio A, Zhang J, Pitkanen J-P, van den Brink J, [26 others], Workman CT, Zamboni N, Nielsen J (*submitted*) Experimental systems biology: Lessons from an integrated, multi-laboratory study in yeast

Nikerel IE, **Canelas AB**, Jol S, Verheijen PJT, Heijnen JJ (*submitted*) Construction of kinetic models for metabolic reaction networks: lessons learned in analyzing short-term stimulus response data

Canelas AB, Ras C, ten Pierick A, van Gulik WM, Heijnen JJ (*submitted*) Finding simplicity in the midst of complexity: a data-driven thermodynamics-based framework for classification and quantification of *in vivo* enzyme kinetics

Zhu W, **Canelas AB**, Landfried K, Stevens A, Holler E, Oefner PJ (*manuscript in preparation*) Quantitative profiling of tryptophan metabolites in human serum and urine by liquid chromatography-tandem mass spectrometry

Oral presentations at international meetings and courses:

Canelas AB et al, Vertical Genomics: Fast dynamic response of central metabolism and glycolytic genes to glucose availability in *S. cerevisiae*, Int. Specialized Symp. on Yeasts XXV (Finland, 2006)

Canelas AB et al, Getting the right numbers: how to avoid some (common) mistakes in metabolomics-based research in *S. cerevisiae*, Yeasterday 2008 (Netherlands, 2008)

Canelas AB et al, Tackling compartmentation: Estimation of the cytosolic free NAD/NADH ratio in yeast using a heterologous indicator reaction, Metabolic Eng. VII (Mexico, 2008)

Canelas AB et al, Getting the right numbers: how to avoid some (common) mistakes in metabolomics-based research in *S. cerevisiae*, Metabolomics Society Conf. IV (USA, 2008)

Canelas AB et al, *In vivo* thermodynamic and kinetic modelling of the central metabolism of microbial production hosts, World Cong. on Industrial Biotechnology and Bioprocessing VI (Canada, 2009)

Canelas AB et al, Three Lessons on Yeast Metabolomics: Method Development, Workflow Validation and Application to Kinetic/Thermodynamic Modelling, KU Leuven Summer School on Metabolomics and Metabolic Networks (Belgium, 2009)

Canelas AB et al, Finding simplicity in the midst of complexity: a data-driven thermodynamics-based framework for classification and quantification of *in vivo* enzyme kinetics, TU Delft Advanced Course on Metabolomics for Microbial Systems Biology (*scheduled, 2010*)

Conference proceedings:

Nikerel IE, **Canelas, AB**, Jol S, Verheijen P, Heijnen JJ (2009) Construction of Kinetic Models for Metabolic Reaction Networks: Lessons Learned in Analyzing Short Term Stimulus Response Data, Mathmod 2009

Bouwman J, van Eunen K, Tuzun I, Postmus J, **Canelas AB**, van der Brink J, Lindenbergh PA, Teixeira De Mattos MJ, Smits GJ, Brul S, Hellingwerf KJ, Westerhoff HV, Bakker BM (2006) Standardization and *in vivo*-Like Enzyme Activity Measurements in Yeast, 2nd Int. ESCEC Symposium on Experimental Standard Conditions on Enzyme Characterizations

Curriculum vitae

André was born on July 10th 1980 in Lisbon, Portugal. He grew up in the area of Oeiras/Cascais, in the coastal outskirts of Lisbon. He did most of his studies until highschool at Escola Salesiana de St. António do Estoril. He spent much of his Summers with his godfather's family in Utrera, in the south of Spain.

As a child he wanted to be an astronaut. As a teenager he wanted to become an Aerospace or Naval Engineer. During highschool (in the mid-90s) he got interested in Genetics and thus went on to study Biological Engineering at Instituto Superior Técnico. After a final-year internship at Bayer (in Wuppertal, Germany) he concluded that he should pursue a PhD. He began working at the Department of Biotchology of T.U. Delft a few months later. This thesis describes the main results of his work on quantitative analysis of *in vivo* kinetics and metabolic networks in *S. cerevisiae*.

Acknowledgements

A great many people have contributed in one way or another to the work that I have carried out in the last 6 years, both that which is presented in this thesis and that which is not. This includes not only all the staff and members of the BPT group (over the years…), but also my partners in the VG project, colleagues and staff of the Kluyver, especially the IMB group, collaborators in the YSBN project, etc. One of the aspects I most appreciated about working in the Kluyver was the possibility to interact with people of different experience and expertise and the freedom to, when faced with a problem, walk to any other office and ask someone their opinion or assistance. I always felt welcome and could usually find an answer, or at least a new direction. The same holds true for my experience with the partners in the VG consortium, among whom I always felt that my input was welcome, and on whom I could count for critical feedback. For this, I am truly indebted to all of you. Yet, tradition holds that at this point I should make an effort to mention all those I am thankful to. Thus, despite recognizing the inherent danger of missing someone, here's a modest attempt to do so.

I shall begin by thanking Sef Heijnen and Walter van Gulik, for deciding to hire me back in the beginning of 2004. You will be forever remembered for your knowledgeable guidance, for your trust in me, at the onset, throughout the project and later when we decided for an extension, for always putting me at ease to speak openly, for your continued availability to discuss matters despite my soon-well-established record of time overruns at meetings and not-so-short drop-ins (I reckon we must have spent more time in discussions outside scheduled meetings than within), and for your loose style of supervision, which although at times seemed frustratingly hands-off, eventually paid off in terms of the learning experience. And, of course, for all the things I've learned from you. Thank you. I would also like to acknowledge the rest of the scientific staff of the group who, although not formally responsible for my supervision, nevertheless kindly assisted me throughout the project. I would like to thank Wouter van Winden for his encouragement, positive attitude and constructive feedback (I shall remember you, among other things, for a peculiar off-the-shelf mixing equipment you brought to the lab for me once…), Peter Verheijen for not giving up educating me in the benefits of proper statistics, many interesting discussions and general good advice, and Aljoscha Wahl, for insightful suggestions and critically reviewing some of the chapters. Walter and Peter are also acknowledged for translating from English the Dutch portions of this thesis.

I am also indebted to the analytical team, not only for analyzing so many of my samples, which often required well-beyond-routine procedures and multiple re-checks, but also for all the development and validation work, which allowed more metabolites to be measured, and for the great two-way interaction, including teaching me so much about MS. I shall remember Cor Ras as one of the most dedicated professionals I've worked with (how many times did we meet in the bike shed, me going out, late as usual, and you coming back in?), Angela ten Pierick for her persistence and attention to detail, Reza Seifar for his patience and diligence (often he

wouldn't have an answer to my stranger questions right away, but he wouldn't let go and typically he'd come back with the solution a couple of hours later, having searched in his books and papers), Jan van Dam for his support and encouragement, especially in the first years, and Zhen Zeng for her contribution in the later years. Johan Knoll also contributed to the analysis effort at some point, and was always a cheerful presence in the corridor, as was Stef van Hateren. Max Zomerdijk also contributed with GC measurements. Still in the topic of analytics, I am indebted to Chiara Cipollina for her work with Angie on the method for determination of PPP metabolites, which was essential in enabling the studies that became chapter 6, and thankful to Nanette Boyle for our collaboration during her visit. Lastly, I am also very indebted to the staff in the fermentation lab, namely Rob Kerste and Dirk Geerts for their reliable and knowledgeable support whenever there were glitches with the equipment or changes to the setups were needed, as well as Mario Pronk and Susan van den Bulk, who joined/left the FTD over the years.

A special word of appreciation goes out to my two closest collaborators, who are also my two paranymphs. Emrah Nikerel has been a roommate and a neighbor, as well as a colleague and a friend. I guess I will remember him especially for his many quirks (his "dynamic" appearance, Godfather, Hat-trick, etc), but I am also indebted for his support, for the exchange of ideas on the directions of our projects and for never failing to satisfy my questions in the subject of modeling. Joost van den Brink has also been both a colleague and a friend. I think he will be remembered mostly for his generous laugh and his enthusiasm for research, and I am also grateful for our many discussions (often on our way to/from project meetings), positive input and our successful collaboration. Furthermore, I am indebted to both Joost and Emrah in a very concrete way, because what I learned during our collaborations played a major role in setting my mind into the direction that resulted in Chapter 6.

I am also grateful to all the other members of the BPT group, for all the experiences we shared. Roeland Costenoble, Penia Kresnowati, Mlawule Mashego and Liang Wu are acknowledged for coaching me in certain specific techniques during my first years. I would also like to thank Sergio Rossel and Domenico Bellomo for their companionship, inspiration and long discussions (both will be remembered for their devotion to Science), Zheng Zhao for good suggestions and teaching me a thing or two about flux analysis, and Michiel van Leeuwen for interesting discussions (it was always a pleasure to serve as bouncing board for some of your ideas and to listen to your exciting descriptions of your micro-scale breakthroughs). I am also very grateful to the students I supervised, Tjerko Kamminga, Stefan Jol, Mehmet Ucisik and Nathalie Bleijie, for the pleasure of sharing their accomplishment and the overall experience of coaching them. In this respect, Emrah, Penia and Sergio are acknowledged for their contribution in co-supervision. Finally, I'd like to thank Hilal Taymaz, Rutger Douma, Lodewijk de Jonge and Marco de Groot for useful discussions and exchange of ideas, and the older (Roelco Kleijn, Uly Nasution, Fredrik Aboka) as well as the newer colleagues (Luísa Cruz, Elaheh Jamalzadeh, Katelijne Bekers, Rob Brooijmans, Amit Deshmukh), and all the students and guests in our group (namely Marilia Foukaraki, Joana Batista, Marc Carnicer, Marjan De Mey), for their continued support and occasional rapid sampling duties.

One could say that my "second allegiance", aside from the BPT group, was to my partners within the Vertical Genomics project, to whom I am grateful for the great opportunities for

collaboration and exchange of know-how. Besides Joost, I would like to thank Jildau Bouwman for her support and efforts in mRNA analysis, Karen van Eunen for interesting discussions and collaboration opportunities, as well as Jarne Postmus, Isil Tuzun and Ronald Aardema for the great interaction over the years. I am also very grateful to Pascale Lapujade for regular discussions and generous encouragement, Barbara Bakker and Gertien Smits for all the active and critical feedback, and Hans Westerhoff, Joost de Mattos, Chris de Koster and Stanley Brul, as well as the industrial partners (Hein Stam, Mike Walsh and John Chapman) and IOP contacts (Tonnie Rijkers, Jan Smeitink, Corine de Rijke, Janneke Timmerman) for some lively discussions during our meetings. Finally, it was through the VG connections that I met Rick Orij, whom I thank for sharing data ahead of publication.

The incorporation of Chapter 5 in this thesis results from my involvement in the inter-lab experiment of the Yeast Systems Biology Network, which provided an interesting opportunity for interaction and collaboration with many other yeast researchers across Europe. In this respect, I would like to thank in particular Jens Nielsen and Dina Petranovic for their continued trust and support, especially when Wouter had to leave the project. And a special word of appreciation goes out to Alessandro Fazio, who came to Delft to assist with the experiments. I shall remember him for his absolutely contagious cheerful attitude and enthusiasm. The period we spent carrying out the experiments in Delft was truly a change from routine. I would also like to thank all the colleagues in each of the partner laboratories for their efforts and contributions, and in particular to Nicola Zamboni and Uwe Sauer for positive feedback and encouragement, as well as to the BPT colleagues who kindly assisted with the samplings.

If the VG team was my "second allegiance", the Industrial Microbiology group felt, at times, almost like a "second affiliation", for the kind and welcoming way in which they always treated me. Besides Pascale, I am very grateful to Jack Pronk for scientific input and his extremely positive attitude, Ton van Maris for interesting discussions and Han de Winde, and Jean-Marc Daran for occasional contributions. I am also thankful to Erik Hulster and Marijke Luttik for knowledgeable technical advice and support, and to Jasper Diderich, Lucie Hazelwood, Leonie van Dijk and Derek Abbott for the free exchange of ideas and information. Finally, I'd like to thank all the IMB colleagues over the years, including Tânia Veiga, Inês Antunes, Andreas Gombert, Wouter Wisselink, Erwin Suir, Zita van der Krogt, Thiago Basso, Irina Bolat, Victor Medina, Barbara Crimi, Stefan de Kok, Eline Huisjes, Rintze Zelle, Eleonora Bellissimi, Diana Harris, Jan-Maarten Geertman, Marinka Almering, Ishtar Snoek, Maurice Toirkens and Siew-Leng Tai, as well as the "BIRDies" Marko Kuijper and Tom Schuurman, for the great interaction and always making me feel at home in "the other side".

Overall, I would also like to thank all the scientific staff and PhD colleagues from the other research groups for their willingness to share their know-how and expertise when I came to them with my questions or problems. I am particularly grateful to Dick de Ridder (EWI) and Hans van Dijken for useful tips, Peter Verhaert for interesting discussions and accepting to be my reservelid, Arthur Kroon for enthusiastic discussions and Rosario Franco and Ron Winkler (BIRD) for molecular biology assistance. And I would also like to say thanks to the organizers of the Kluyver Colloquia (Wouter van Winden, Marcel Ottens, Peter-Leon Hagedoorn,

Cristian Picioreanu) and Advanced Courses (Lies van der Meer-Lerk and Ger Aggenbach) for the opportunity to listen to many excellent presentations.

Furthermore, I would like to acknowledge all the staff of the Kluyver for providing and caring for the infrastructure that was essential to making these studies possible. I would especially like to thank Sjaak Lispet for his staggering ability to juggle his many functions effectively, Astrid van Uijen, Apilena Sapioper, Jos Lispet, Herman Frumau and Leslie Robertson for their excellent support to the fermentation laboratories, Hans Kemper (and Marcel van den Broek in the early years) for the knowledgeable IT support, Arno van den Berg and the workshop team for their valuable creations, and the entire organization, management and finances team for keeping everything in order, in particular Carla Segaar, Annette Scharp and Jenifer Baptiste for their kindness and ready assistance, and Wim Morien, who will be remembered for his patience and helping with some of my first "baby steps" into the Dutch language. Finally, I'd also like to thank the whole Botanical Garden team for keeping the garden beautiful and letting us use it when it was sunny outside.

I think that takes care of the more tangible contributions to the work, so to close up I would like to thank all of those who made a "soft" contribution just by being part of my life in Delft for the last 6 years, in as well as outside the lab, for your friendship and company, for the fun we had in the good times, and for the support in the troubled times. I would like to thank in particular all the "regulars" at the BPT/BST coffee room, with whom I shared so many meals and long discussions, especially Domenico, Rutger, Esteban Freydell and Çağrı Efe for the sharing of experiences and expectations, as well as Margarida Temudo and João Xavier, my daily lunch companions at the Aula in the early years, for their company and advice on life and science (Margarida will be remembered for her hearty laugh and her determination, while João still comes to mind for his scientific drive and his metaphorical one-liner: The first year is just "pitching" the stone). I will also hold fond memories of the great times spent in the company of the Football For Nerds gentlemen and the Flush It In gamblers, of Parag Shanbhag, Xhemile Bakiu&Peter, Allard Friedrich, Miguel Mendes and the rest of the Portuguese/Spanish/Italian Erasmus gang I met at the "RolaHola" in my first year, and of the many friends and acquaintances made over the years in the social circles in and around Delft (namely, besides many of the aforementioned, Leila Dias, Laura Araújo, Filippo& Orsi Rosatti, Anish& Zsófi Patil, André&Katrin Bardow, Mike&Marlies Reiha, Marta&Paulo Neves, Amir Goudarzi, Inés Alvarez&Willem, Eva Sanchez, Carol Engel&Diego and Pedro Pereira). On the subject of parties, I would also like to acknowledge the people with whom I organized some events at the Kluyver, namely Penia (BPT karting uitje), Janine Kiers (lab uitje), Emrah (BPT paintball uitje) and Rintze (Darwin pub quiz). Finally, a special thanks goes out to Paulo Moreno, friend and comrade for nearly 15 years, for keeping the good times rolling.

Last but not least, I would like to acknowledge the two most important individuals in my life: Fanni, the sweetest woman I ever met, for leaving her family and friends to join my adventures in Holland, and for putting up with my stress and frequent absence, especially in the later stages; and my mother, the most generous and dedicated person I know, for her continuous support, advice and encouragement and, ultimately, for having made all of this possible.

All of [science] is either impossible or trivial.
It is impossible until you understand it, and then it becomes trivial.

Ernest Rutherford


extraction
concentrations
used
method
intracellular
methanol
quenching
metabolic
measured
data
cell
model
enzyme
rate
results
determined
in-vivo
changes
experiments
quantitative
study
higher
increased
complexity
cytosolic
glucose
compared
values
possible
cerevisiae
represent
large
free
effects
analyzed
analytical
application
considered
indicate
fraction
protocol
metabolomics
constant
number
NAD
NADH
organic
fact
cyt
time
set
procedure
specific
generated
described
development
expression
parameters
comparison
cultures
solution
addition
experimental
leakage
provide
techniques
treatment
performed
literature
example
relative
process
categories
weight
extent
occur
relevant
according
ethanol
applied
approximately
platforms remain
quantification
recent
accessible
Discussion
role
duplicate
glycolysis
project
Genomics
abundance
FTM
central
pulse
allow
obtained
require
format
acetaldehyde
regarding
step
near-equilibrium
Finally
dataset
turnover
close
Mtl1P
PEP
balance
adequate
biomass
Chapter
regulation
point
averages
reduction
high
key
glycerol
practice
ATP
G6P
MIPDH
essential
biosynthesis
principle
make
molecular
μmol
implies
previous
large-scale
absence
place
varied
multiple
release
carbon
best
chemical
thesis
peptide
defined
aspect
stages
HW
ΔG
gave
supply
LC-MS
thought
comprising
min
alternative
deviation
perturbation
physiological
prediction
confirmed
proposed
improved
apparent
phosphorylated
classes
assumption
expected
challenges
including
description
substantial
correlation
itself
amino-acids
substrate
mass
linear
coli
aerobic
control
effectors
differential
inactivation
filtration
coenzyme
followed
variations
important
centrifugation
particularly
Instead
significantly
comprehensive
exposure
identified
order
polar
task
way
methodologies
addressed
compartments
integrated
PEP
mannitol
short-term
tube
purpose
YSBN2
phosphate
water
operate
energy
redox
taken
proteomics
state
F6P
systematic
AANM
given
gDW
direction
biological
complete
known
wide
highly
distinguish
mM
3PG
CM
FBP
broth
aimed
targeted
reduced
spiking
main
M6P
T6P
MS
good
matrix
pellets
β
work
detailed
assumed
depends
suggests
current
simultaneously
nucleotides
capacity
agreement
Succinate
question
eukaryotes
overall
calculated
investigated
transcription
relation
internal
conversion
Pi
cultivation
reason
Tested
derived
conventional
Leu
evaluated
demonstrate
affect
decrease
fermentation
shall
choice
13C-labeled
in-vitro
relies
labs
cause
sufficient
internal
sensor
excess
Boiling
contrast
come
versus
far
clear
energy
introduction
filter
assay
actually
Malate
occurrence
form
smaller
sensitive
play
advantage
absolute
uncertainty
highlight
subsequent
objective
somewhat
vortexed
adapted
outcomes
proportion
driving
reflect
degrees
μl
mol
solvent
whole-broth
goal
tools
microbial
Citrate
especially
extensive
maximum
optimal
overview
dilution
synthesis
collected
understand
introduced
independent
immediately
industrial
Ultimately
magnitude
potential
Additionally
pyruvate
grown
precise
Oxoglutarate
2PG
underestimation
strategy
content
enabling
analyses
shows
attempted
pre-cooled
high-quality
mixed
trace
period
stored
appear
oxygen
feasible
amongst
T-EX
nearly
coverage
vs
gap
ionic
ENO
strength
biochemical
constitute
partly
affinity
treated
ratio
mtlD
NAD-NADHxNxDpH