

Deciphering population dynamics as a key for process optimization

DISSERTATION

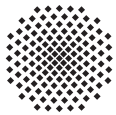
Von der Fakultät Energie-, Verfahrens- und Biotechnik der Universität
Stuttgart zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)
genehmigte Abhandlung

Vorgelegt von

Sarah Lieder

aus Rotenburg (Wümme)

Hauptberichter: Prof. Dr. Ralf Takors
Mitberichter: Prof. Dr. Han de Winde
Tag der mündlichen Prüfung: 25.09.2015



University of Stuttgart
Germany

Institut für Bioverfahrenstechnik

2016

DECLARATION

I declare that the submitted work has been completed by me and that I have not used any other than permitted reference sources or materials. All references and other sources used by me have been appropriately acknowledged in the work.

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Es wurden von mir nur die in der Arbeit ausdrücklich benannten Quellen und Hilfsmittel benutzt. Übernommenes Gedankengut wurde von mir als solches kenntlich gemacht.

San Francisco, den 20.08.2016



Sarah Lieder

 CONTENTS

List of Figures	vii
List of Tables	ix
Glossary	xi
Zusammenfassung	xv
Summary	xvii
1. Motivation and objectives	1
1.1. Heterogeneity in microbial cultivations	2
1.2. <i>P. putida</i> KT2440 as a promising industrial production host	5
2. <i>Pseudomonads</i> as organisms of interest	7
2.1. <i>P. putida</i> as industrial production host	8
3. Characterization of microbial populations	11
3.1. The origin of population heterogeneity in clonal bacterial populations	12
3.2. Cultivation strategies and experimental methods for deciphering population dynamics	15
3.2.1. The chemostat as a model system to decipher population dynamics	15
3.2.2. Experimental and analytical methods for describing cell populations and distributions of cell properties	17
4. Modeling microbial populations	21
5. Material and Methods	25
5.1. Bacterial strains, media and cultivation systems	25
5.2. Nucleic acid manipulation and plasmid construction	26
5.3. Analytical methods	28
5.4. Flow cytometry analysis, cell sorting and subpopulation-proteomics	28

5.5.	Transcriptome analysis	29
5.5.1.	Sampling procedure and RNA next generation sequencing	29
5.5.2.	Statistical data analysis	30
5.6.	Quantification of cultivations	31
5.6.1.	Bacterial growth kinetics	31
5.6.2.	Mass balances	32
5.6.3.	Carbon balancing	36
5.6.4.	Maintenance demands	36
5.6.5.	Propagation of uncertainties	36
6.	The cell cycle as origin of population dynamics	37
6.1.	Design of the experimental set-up	38
6.2.	Physiological characterization of the average population	39
6.3.	Quantification of subpopulations via flow cytometry	39
6.4.	Subpopulation proteome analysis	41
7.	The environmental condition as origin of population dynamics	47
7.1.	Design of the experimental set-up	48
7.2.	Quantification of subpopulations via flow cytometry	48
7.3.	Quantification of stress impact on population dynamics using mathematical modeling	51
7.3.1.	Mathematical framework	51
7.3.2.	Implementation of the mathematical model	52
7.3.3.	Quantitative impact of stress on cell cycle phases	54
8.	Optimizing microbial cell chassis by streamlining the genome	61
8.1.	Reaction parameters and energy profile of streamlined-genome derivatives of <i>P. putida</i> KT2440	62
8.2.	Heterologous protein synthesis in streamlined-genome derivatives of <i>P. putida</i> KT2440	65
9.	Conclusions and perspectives	69
	Author contributions	75
	Acknowledgements	77
	Curriculum vitae	79
	References	81
	Appendices	99
A.	Manuscript I	99
A.1.	Introduction	100
A.2.	Materials and Methods	102
A.3.	Results	105
A.4.	Discussion	110
A.5.	Acknowledgments	111
A.6.	Supplemental material	112

B.	Manuscript II	114
B.1.	Introduction	114
B.2.	Material and Methods	116
B.3.	Results	119
B.4.	Discussion	124
B.5.	Acknowledgments	126
B.6.	Supplemental material	127
C.	Manuscript III	132
C.1.	Introduction	132
C.2.	Material and Methods	135
C.3.	Results and Discussion	141
C.4.	Conclusion	151
C.5.	Acknowledgements	153
C.6.	Supplemental material	154

LIST OF FIGURES

1.1.	Summary of biological and technical factors contributing to a robust production process.	2
3.1.	Schematic description of how average measurements mask real population states.	12
3.2.	Overview of origins of population heterogeneity.	13
3.3.	Schematic overview of the bacterial cell cycle.	14
3.4.	Schematic chemostat set-up.	16
3.5.	Schematic flow cytometer set-up.	18
4.1.	Classification of mathematical models used to describe microbial cultivations.	22
5.1.	Schematic working cell bank procedure	26
5.2.	Schematic 3.7L bench-top reactor set-up.	27
5.3.	Schematic overview of the bacterial growth curve.	32
6.1.	Physiological data of <i>P. putida</i> KT2440 continuous cultivations at different growth rates.	38
6.2.	DAPI fluorescence (DAPI) and forward scatter (FSC) as parameters for cell sorting.	40
6.3.	Subpopulation distributions at different growth rates.	41
6.4.	Visualization of COG annotation.	42
6.5.	Circular treemaps visualizing differentially expressed functional protein categories.	44
7.1.	Overview of the experimental set-up to investigate the impact of environmental stress on population dynamics.	50
7.2.	Illustration of the calculation of DNA histograms $n(G)$	53
7.3.	Duration of the replication phase in dependence of the specific growth rate μ .	56
7.4.	Duration of cell cycle phases in dependence of the specific stress condition.	57
8.1.	Design of reduced-genome <i>P. putida</i> KT2440 strains.	62
8.2.	Growth parameters of <i>P. putida</i> KT2440, EM329 and EM383 in glucose-limited chemostat cultures.	63
8.3.	Summary of energy parameters of <i>P. putida</i> KT2440, EM329 and EM383 in glucose-limited continuous cultures.	65

8.4.	Impact of <i>gfp</i> expression on the maximum specific growth rates of <i>P. putida</i> KT2440, EM329 and EM383.	67
8.5.	Heterologous protein production in <i>P. putida</i> KT2440/pS234G, EM329/pS234G and EM383/pS234G	68
A.1.	Schematic overview of the bacterial cell cycle.	100
A.2.	Summary of the physiological state of the average population.	105
A.3.	Dot plots of DNA content versus forward scatter at different growth rates 0.1 h^{-1} , 0.2 h^{-1} and 0.7 h^{-1}	106
A.4.	Circular treemaps visualizing differentially expressed functional protein categories.	108
A.5.	Heatmaps of metabolic pathways of special interest.	109
A.6.	Replicate dataset of dot plots of DNA content versus forward scatter at different growth rates 0.1 h^{-1} , 0.2 h^{-1} and 0.7 h^{-1}	112
A.7.	Overview of the total protein detection and protein annotation.	113
B.1.	Overview of the experimental set-up (a) and the workflow of data-based modeling (b).	120
B.2.	Durations of the replication phase in dependence of the specific growth rate μ	122
B.3.	Durations of cell cycle phases in dependence of the respective stress condition	123
B.4.	Illustration of the calculation of DNA histograms $n(G)$	130
B.5.	Summary of the time course of the solvent stress chemostat.	131
C.1.	Rationale behind the design of reduced-genome derivatives of <i>P. putida</i> KT2440.	134
C.2.	Summary of the growth parameters for the different strains under study in glucose-limited chemostat cultures.	142
C.3.	Characterization of energy parameters for the different strains under study in glucose-limited chemostat cultures.	145
C.4.	Flow cytometry analysis of the green fluorescent protein accumulation in the strains under study.	147
C.5.	Characterization of growth parameters and protein production kinetics for the different strains under study in batch bioreactor cultures.	149
C.6.	Physiological characterization of (A) <i>P. putida</i> KT2440, (B) <i>P. putida</i> EM329, and (C) <i>P. putida</i> EM383 in glucose-limited chemostat cultures at different dilution rates (D).	154
C.7.	Carbon balance of glucose-limited chemostat cultures of <i>P. putida</i> KT2440, <i>P. putida</i> EM329, and <i>P. putida</i> EM383.	154
C.8.	Propidium iodide (PI) exclusion was used to estimate cell viability in <i>P. putida</i> KT2440, <i>P. putida</i> EM329, and <i>P. putida</i> EM383 with the empty and the recombinant plasmid.	155
C.9.	Physiological characterization in bioreactor batch cultivations of the different strains carrying plasmids.	156
C.10.	Physiological characterization in bioreactor batch cultivations of the different strains carrying plasmids.	157

 LIST OF TABLES

7.1. Population composition at non-stressed and stressed conditions analyzed by flow cytometry	50
7.2. Summary of the duration of cell cycle phases and goodness of fit of the simulation. . .	55
7.3. Differentially expressed genes under decanol stress conditions, annotated in the functional group ‘replication’	58
B.1. Summary of the duration of cell cycle phases and goodness of fit of the simulation. . .	121
B.2. Differentially expressed genes under decanol stress conditions, annotated in the functional group ‘replication’	124
B.3. Summary of the duration of cell cycle phases and goodness of fit of the simulation. . .	131
C.1. Bacterial strains and plasmids used in this study	136
C.2. Growth and protein synthesis parameters in shaken-flask cultures of different recombinant <i>P. putida</i> strains	146

GLOSSARY

a	normalized age of a cell within one generation time τ
AEC	adenylate energy charge
A.F.U.	arbitrary fluorescence units
$b(\mathbf{y}, t)$	breakage function; it represents the probability of cell division at the physiological state \mathbf{y} as a part of a dynamic population balance equation
C1	subpopulation with a single chromosome, representing cells in B phase that just divided and did not start to replicate their DNA yet
C2	subpopulation C2 containing two chromosomes, representing cells in pre-D or D phase that finished replication but did not divide yet
CDW	biomass cell dry weight (in gL^{-1})
CER	carbon dioxide evolution rate (in $molL^{-1}h^{-1}$)
c_{O_2}	oxygen concentration in the bulk liquid (in $molL^{-1}$)
$c_{O_2}^*$	oxygen concentration in the liquid which is in equilibrium with the gas phase (gas-liquid interphase) (in $molL^{-1}$)
CO_2	carbon dioxide
COG	clusters of orthologous groups (database)
Cx	subpopulation with more than doubled chromosome content, representing cells performing multifork DNA replication
D	Dilution rate (in h^{-1})
DAPI	4',6-diamidino-2-phenylindole (fluorescence stain used in Flow Cytometry)
F	flow rate (in h^{-1})
FACS	Fluorescence Activated Cell Sorting
FCS	functional class scoring (gene set analysis method)
FS	forward scatter (measurement in Flow Cytometry)
$f(\mathbf{y}, t)$	distribution function; it describes the distribution of cells characterized by the internal state vector \mathbf{y} at time t
$G(a)$	cellular DNA accumulation function
GFP	green fluorescent protein

GO	Gene Ontology
GSA	Gene Set Analysis
HPLC	high pressure liquid chromatography
KEGG	Kyoto Encyclopaedia of Genes and Genomes
k_{La}	mass transfer coefficient
K_s	limiting substrate concentration at which the specific growth rate is half its maximum value
LFQ	label-free quantification
m_s	maintenance coefficient (in $\text{g}_{\text{GLC}}\text{g}_{\text{CDW}}^{-1}\text{h}^{-1}$)
$n(a)$	represents the probability density of a single cell in a population to have a certain age a
$n(G)$	theoretical DNA distribution
$n(t)$	quantity of cells in the population per property space V_y
OD	optical density
ORA	over-representation analysis (gene set analysis method)
OTR	oxygen transfer rate (in $\text{molL}^{-1}\text{h}^{-1}$)
OUR	oxygen uptake rate (in $\text{molL}^{-1}\text{h}^{-1}$)
p	pressure (in Pa)
PBE	Population Balance Equation
PBM	Population Balance Model
PHA	polyhydroxyalkanoates
$p\text{O}_2$	dissolved oxygen partial pressure
$p(\mathbf{y}, \mathbf{y}^*, t)$	partitioning function; it specifies the probability of a cell at state \mathbf{y}^* to divide into two daughter cells with the physiological states \mathbf{y} and $\mathbf{y}^* - \mathbf{y}$ as a part of a dynamic population balance equation
q_p	biomass specific production rate (in $\text{gg}^{-1}\text{L}^{-1}$)
q_s	biomass specific substrate uptake rate (in $\text{gg}^{-1}\text{L}^{-1}$)
R	universal gas constant (in $\text{Jmol}^{-1}\text{K}^{-1}$)
r_c	maximum replication rate (in kbp/min)
RQ	respiratory quotient
$r_s(\mathbf{y}, s)$	volumetric substrate consumption rate (in $\text{gL}^{-1}\text{h}^{-1}$)
$r(\mathbf{y}, t)$	single cell growth rate; it describes the rate of accumulation of a property within a cell as a part of a dynamic population balance equation
S	substrate concentration (in gL^{-1})
SG	streamlined-genome
SS	side scatter (measurement in Flow Cytometry)

t	time (in h)
T	absolute temperature (in K)
TCA	tricarboxylic acid cycle
t_d	doubling time in h
TP	pathway topology based gene set analysis
V	cultivation volume (in L)
\dot{V}_g	volumetric gas flow rate (in Lh^{-1})
V_y	volume of the property space with cells being described by the internal state vector \mathbf{y} at time t
X	biomass concentration (in gL^{-1})
y_i	volumetric gas fraction of gas i (in %)
$Y_{P/X}$	product per biomass yield coefficients (in gg^{-1})
$Y_{X/S}$	biomass per substrate yield coefficients (in gg^{-1})
$Y_{X/S_{true}}$	true yield of biomass on glucose (in $g_{CDW}g_{GLC}^{-1}$)
μ	specific growth rate (in h^{-1})
τ	generation time (in h)

ZUSAMMENFASSUNG

Die Biotechnologie ist einer der am stärksten wachsenden Wirtschaftszweige des 21. Jahrhunderts und ermöglicht die nachhaltige Produktion von industriell bedeutenden Verbindungen mit Hilfe von Zellfabriken. Um mit klassischen chemischen Produktionsprozessen wirtschaftlich konkurrieren zu können, müssen sowohl biologische Kenntnisse über beteiligte Stoffwechselwege, die Zellfabrik und deren Populationsdynamik unter harschen Produktionsbedingungen, als auch Wissen aus Ingenieursdisziplinen, wie der Bioprozesstechnik und dem Bioreaktordesign, berücksichtigt werden.

Bis heute wird bei der Charakterisierung biologischer Prozesse ein uniformes Zellverhalten angenommen. In den letzten Jahren wurde jedoch bekannt, dass isogene mikrobielle Populationen aus Zellen mit unterschiedlichen Phänotypen bestehen, die einerseits zu Leistungsverlust durch weniger produktive Subpopulation führen können, aber andererseits durch schnellere Anpassungsfähigkeit der Population zur Robustheit des Prozesses beitragen. Um die Entstehung von Populationsheterogenitäten in Bioprozessen kontrollieren zu können, müssen die zugrunde liegenden Mechanismen verstanden werden. Im ersten Teil der Arbeit wurde der Einfluss des Zellzyklus und industriell relevanter Stressbedingungen auf die Entstehung von Populationsdynamiken quantifiziert.

Zunächst wurde die Abhängigkeit der Proteinzusammensetzung von *Pseudomonas putida* KT2440 Zellen in unterschiedlichen Zellzyklusphasen unter langsamen und schnellen Wachstumsbedingungen untersucht. Überraschenderweise konnten keine signifikanten Unterschiede in dem Proteom der durch Flusszytometrie detektierten Zellzyklus-Subpopulationen festgestellt werden. Im Gegensatz dazu verursachte die Veränderung der Wachstumsrate große Unterschiede in der Proteinzusammensetzung z.B. in Bezug auf Kohlenstoffspeicherung, Motilität und der Translationsmaschinerie. Die Ergebnisse zeigen, dass der Zellzyklus selbst einen nur geringen Einfluss auf die Entstehung von Heterogenitäten auf Proteinebene unter den getesteten Wachstumsbedingungen hat, während die Wachstumsrate die Proteinzusammensetzung klar bestimmt.

In großvolumigen Prozessen werden Zellen anspruchsvollen und sich ständig ändernden Mikroumgebungen ausgesetzt. Unter verringerter Eisen- und Sauerstoffverfügbarkeit und Lösungsmittelexposition –typischen Stressbedingungen, denen *P. putida* Zellen in industriellen Anwendungen begegnen– wurde eine Veränderung der Populationszusammensetzung entdeckt, die eine Anpassung der Zellzyklusdynamik innerhalb der Population vermuten ließ. Daraufhin wurde ein datengetriebener Modellansatz verwendet, um eventuelle Unterschiede in den Zellzyklusphasen zu quantifizieren: Bei gleichbleibender Generationszeit verkürzten die Zellen ihre Replikationsphase unter allen getesteten Stressbedingungen. Dem entsprechend verlängerten sich die übrigen Phasen des Zellzyklus: die Zeit zwischen Geburt der Zelle und Start der Replikation und die Zeit zwischen dem Ende der Replikation und der Teilung der Zelle. Die Beschleunigung der Replikationsrate (bis zu 1.9fach) korreliert hierbei mit der Stressbelastung. Transkriptomdaten untermauern die Beobachtungen und zeigen eine Überexpression von Genen, die Komponenten der zellulären Replikationsmaschinerie kodieren und die somit eine Erhöhung der Replikationsgeschwindigkeit erreichen könnten. Unsere Ergebnisse zeigen, dass Zellen unter Stress die Replikation der genetischen Information beschleunigen. Dieses Phänomen ist begleitet von einer ausgewogenen Veränderung der Dauer aller Zellzyklusphasen und dient der Aufrechterhaltung einer konstanten Wachstumsrate.

Im zweiten Teil dieser Arbeit wurde das Potential von gezielter Genomreduktion als Strategie zur Stammoptimierung am Beispiel von *P. putida* Stämmen erörtert. Bei den zuvor entfernten zellulären Funktionen handelte es sich zum Einen um die Flagellamotilität und zum Anderen um Gene, die bei Deletion mit der Verbesserung geno- und phänotypischer Stabilität in Verbindung gebracht werden. Die beiden Genom-reduzierten Mutanten wurden hinsichtlich industriell-relevanter physiologischer Merkmale untersucht und übertrafen den Wildtyp KT2440 in allen untersuchten Merkmalen: Energetische Parameter, wie die Energieladung und der Adenosintriphosphatgehalt der Zellen, waren signifikant erhöht. Zudem benötigten die Mutanten einen geringeren Anteil der Substrataufnahmerate für Erhaltungsstoffwechselprozesse. Desweiteren zeigten die Mutanten verbesserte Biomasserträge und erreichten höhere Wachstumsraten in Batch-Kultivierungen als *P. putida* KT2440. Abschließend wurde die Produktionskapazität von heterologen Proteinen am Beispiel des grün fluoreszierenden Proteins bestimmt: Auch hier zeigten die Mutanten eine Erhöhung der Ausbeute an rekombinantem Protein pro Biomasse von bis zu 40 %. Die Ergebnisse bestätigen, dass mit gezielter Genomreduktion eine Optimierung des Energiehaushaltes und der Produktionskapazität von mikrobiellen Zellfabriken erreicht werden kann.

Zusammenfassend kann der nicht-pathogene Stamm *P. putida* KT2440 als Zellfabrik empfohlen werden. Der Stamm besitzt viele vorteilhafte Eigenschaften für biotechnologische Anwendungen, wie z.B. ein hohes Maß an Stressrobustheit und Stoffwechselvielfalt, eine relativ einfache genetische Manipulierbarkeit und einen GRAS-Status (generally recognized as safe). Eine Kombination aus weiterführenden Optimierungen des Produktionsstammes und einem tieferen Verständnis der zugrunde liegenden Mechanismen von Populationsdynamiken wird in Zukunft mit Sicherheit die Produktionsleistung einer Vielzahl an biotechnologischen Prozessen erhöhen.

SUMMARY

The field of biotechnology forms the foundation for one of the biggest growing industries in the 21st century. The exploitation of cell factories combines the production of valuable compounds with core values, like environmental friendliness and sustainability. However, the feasibility of a biotechnological process stands or falls with its ability to compete economically with the classical chemical production process. In order to design a viable large-scale microbial production process, biological knowledge about the metabolic pathways involved, the cell factory and its population behavior in industrially relevant environmental conditions have to go hand in hand with classical engineering disciplines comprising bioreactor design and bioprocess control. Here, we assessed *Pseudomonas putida* KT2440 as a promising cell factory and focused on the elucidation of its population dynamics as a key for process optimization.

While optimizing the biological process, uniform cell behavior is assumed, thus, leveling individual to ‘averaged’ cell properties. However, recent research manifested a more differentiated picture: Isogenic microbial cultures comprise subpopulations with dissimilar phenotypes that on one hand may cause performance loss due to less productive subpopulations, but on the other hand increase the robustness of the process as a result of faster population adaptation to challenging environments. The ability to control and harness traits of heterogeneous cell populations relies on a deeper understanding of the underlying mechanisms. Here, we quantify the impact of (1) the cell cycle and (2) industrially-relevant stress conditions on population heterogeneity.

Cell cycling and cell cycle decisions are assumed to play a role in the development of population heterogeneity within clonal populations. We investigated the dependency of the protein inventory of subpopulations in different cell cycle phases under slow and fast growth conditions, using a combination of chemostat cultivations, fluorescence activated cell sorting and mass spectrometry based proteomics. Surprisingly, the protein inventory of subpopulations growing at the same growth rate was highly similar and therefore independent of the cell cycle phase. On the contrary, different

growth rates caused major differences in the proteome with respect to e.g. carbon storage, motility and the translational machinery. The results give rise to the assumption that the cell cycle itself has a minor impact on population heterogeneity under the conditions tested, while the growth rate clearly determines the protein composition.

Industrial large-scale fermentations provide challenging and constantly changing environmental conditions for the microbial cell population. Here, we deciphered population dynamics that result from decreased iron or oxygen availability and solvent exposure –typical stress environments *P. putida* strains are facing in industrial set-ups. While quantifying subpopulation distributions via flow cytometry under non-stressed and stressed conditions in chemostats, we observed adjustments of cell cycle dynamics in the population. Data-driven modeling was applied to quantify changes in the durations of the cell cycle phases. Under all stress conditions tested, the replication phase was shortened, while the time from birth until initiation of replication and the time from end of replication until cell division was prolonged accordingly. Thereby, the increase in replication rate (up to 1.9 fold) was correlated to the severity of the stress imposed. Transcriptome data hint towards overexpression of crucial genes related to the replication machinery to achieve the replication speed up. It seems that fast replication of the genetic information is of high priority under stress conditions, resulting in a balanced altering of the duration of all cell cycle phases as a cellular mechanism to maintain constant growth rates.

In the second part of the thesis, we explored *Pseudomonas putida* KT2440 as a promising microbial cell factory. Production hosts can be designed by (1) rational pathway engineering or (2) removing all elements deemed unnecessary for cellular functions other than replication and self-maintenance in order to improve energy availability for production and genomic stability. Following the latter strategy, we evaluated the impact of targeted genome reduction –particularly the deletion of flagellar motility and genes associated with improving genotypic and phenotypic stability– on industrially-relevant physiological traits. The two *P. putida* derivative strains out-competed the parental strain in every trait assessed: At first, energetic parameters were quantified at different controlled growth rates in continuous cultivations and both strains showed a higher adenosine triphosphate content, adenylate energy charge and decreased maintenance demands than the wild-type strain KT2440. Second, the mutants grew faster and reached higher biomass yields in batch cultivations. Finally, when the production capacity of the green fluorescent protein was assessed in the mutants, an up to 40 % increase of recombinant protein yield was observed.

Summarizing, we advocate the non-pathogenic *P. putida* KT2440 as a cell factory of choice, uniting desirable traits for biotechnological application, such as a high level of stress robustness, metabolic diversity, a relative ease of genetic manipulation and a GRAS status (generally recognized as safe). Finally, a combination of a targeted optimization of the production host and a deeper mechanistic understanding of population dynamics will certainly enhance the overall production performance in diverse biotechnological processes.

CHAPTER 1

MOTIVATION AND OBJECTIVES

The biotechnological industry is one of the biggest growing industries of the 21st century – more than 22 million employees are contributing alone in Europe to its €1.5tn business (Gartland et al., 2013). This rapidly advancing market is fueled by scientific and technological progress throughout a wide spectrum of applications, ranging from medical over agricultural and environmental to industrial biotechnology. The exploitation of cell factories to produce valuable compounds is advancing to a key technology, promising environmental friendliness and sustainability (Sauer et al., 2012). However, the feasibility of a biotechnological process stands or falls with its ability to compete economically with the classical chemical production process (Lee et al., 2012).

In order to design a viable large-scale microbial production process, multiple layers of biological and engineering knowledge needs to be gathered and considered comprehensively (Figure 1.1): Biological knowledge of the molecular mechanisms and the metabolic networks involved, the choice of an optimal cell factory and its population behavior have to go hand in hand with classical engineering disciplines comprising bioreactor design and bioprocess control (Sauer et al., 2012). The success of a microbial production process is dependent on interdisciplinary gain of knowledge and optimization attempts. In this thesis, we focus on the central player within these diverse contributions to a cost-effective production process: the cell factory. The choice and construction of a robust microbial host as well as its population dynamics in industrially-relevant environments will be elucidated.

This study was carried out during a three year period within the ‘European Research Area - Industrial Biotechnology’ project ‘*Pseudomonas* 2.0’. The innate potential of non-pathogenic *Pseudomonas* was exploited by a combination of systems analysis to provide a competitive and beneficial alternative to commonly applied bacterial cell factories, such as *Bacillus*, *Corynebacterium glutamicum* or *Escherichia coli*.

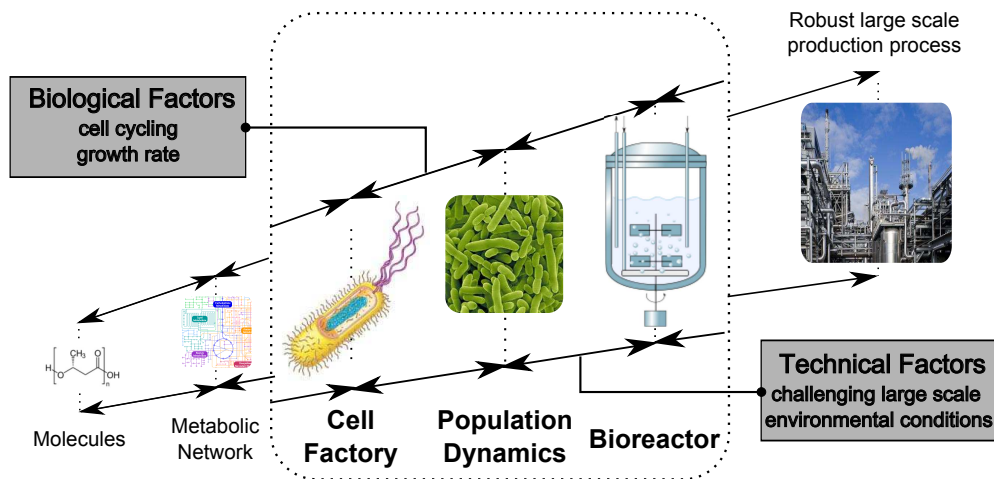


Figure 1.1.: Summary of biological and technical factors contributing to a robust production process. A viable biotechnological process is dependent on the choice of a robust microbial cell factory and its population dynamics, which is governed by an interplay of biological and technical factors. In this thesis, we evaluated the suitability of *Pseudomonas putida* KT2440 as an alternative production host and elucidated its population dynamics in industrially-relevant environments.

Being part of this research consortium, we focused on two closely related research objectives: First, we set out to quantify and investigate the advent and origin of population heterogeneity arising in cultivations of the type strain *P. putida* KT2440 under industrially relevant environmental conditions. Second, we evaluated *P. putida* KT2440 as an efficient cell factory, examining and comparing physiological traits of optimized derivative strains to their parental strain.

The following sections give a brief introduction into the research background and highlight the outstanding questions motivating this thesis.

1.1. Heterogeneity in microbial cultivations

Optimization approaches of fermentation processes are traditionally assuming an uniform isogenic microbial population, thus leveling individual to ‘averaged’ cell properties. However, recent research studies showed a more differentiated picture and revealed that even isogenic microbial cultures comprise individuals that are by no means identical, but exhibit dissimilar phenotypes (Avery, 2006; Nikel et al., 2014c).

Several factors that play a role in the onset of population heterogeneity have been suggested: ‘Internal’ biological factors, such as mutations, gene expression noise or cell cycle decisions, but also ‘external’ technical factors, such as changing micro-environments due to deficient mixing in large scale production, might lead to or further amplify differences in microbial phenotypes (Müller et al., 2010).

On one hand, population heterogeneity is considered as highly unwanted in industrial bioprocesses, because it putatively causes performance loss (Neumeyer et al., 2013). On the other hand, heterogeneity can be beneficial for the robustness of the fermentation process, because it allows faster adaptation of the microbial population to changing environments (Enfors et al., 2001; Hewitt et al., 1999).

Controlling and harnessing traits of heterogeneous cell populations will certainly improve biological production processes, but rely on a deeper understanding of the underlying mechanisms, which, until now, are mostly unknown (Müller et al., 2010; Díaz et al., 2010). In this thesis, two suggested key players in the onset and amplification of population heterogeneity were investigated: the cell cycle and the environmental condition.

The cell cycle as a driver of population heterogeneity

The first research objective covers the investigation of the cell cycle as a biological factor causing population heterogeneity (chapter 6). Cell cycling and cell cycle decisions are assumed to play a key role in the development of population heterogeneity within clonal populations (Avery, 2006; Müller et al., 2010). In the field of applied microbiology, scientists argue whether specific cellular processes occur only in dependency of the cell cycle phase (Mitchison, 1977): Energetic costly processes, e.g. product synthesis, could be accomplished by the cell within the stochastic phases of the cell cycle, where neither replication nor cell division occurs (Bley, 1990; Müller et al., 2010).

Well defined experimental studies are fundamental for investigating the origin of population heterogeneity (Lencastre Fernandes et al., 2011). Here, the following work packages were designed to shed light on the role of the cell cycle as a driver of population heterogeneity:

- Set-up of a reproducible and controlled fermentation process using *P. putida* KT2440, including the development of a seed-train
- Definition of a suitable reference condition and the accomplishment of three characteristic biological replicate cultivations
- Development and test of sampling and sample processing techniques for representative biomass samples suitable for subpopulation and proteome analysis
- Statistical assessment and analysis of flow cytometry datasets
- Definition of a meaningful subpopulation separation variable
- Statistical assessment and analysis of subpopulation proteome datasets
- Comparison of the protein inventory of cell cycle subpopulations

The environment as a driver of population heterogeneity

The second research objective is concerned with deducing the role of the environment as an external factor triggering the advent of population heterogeneity (chapter 7). Besides the above mentioned biological side, technical settings can also cause and amplify population heterogeneity. During large scale industrial fermentations, cells are exposed to less ideal conditions as compared to laboratory scale cultivations: Even though process parameters are tightly controlled, significant local gradients of e.g. dissolved oxygen availability cannot be avoided due to limited mixing and mass transfer (Schweder et al., 1999). Cells circulating through the bioreactor are exposed to shifts in their environment (Fritzsche et al., 2012) and need to continuously adjust their physiology to cope with these fluctuating conditions (Enfors et al., 2001). Regarding the production process itself, it was observed that industrial environments lead to undesired population physiologies including subpopulations of reduced biomass yield and productivity (Lara et al., 2006; Enfors et al., 2001; Carlquist et al., 2012).

Even though population heterogeneity is by now a widely accepted fact, it is rarely taken into account in optimization strategies of bioprocesses (Müller et al., 2010). In order to integrate the interplay of changing environments and subpopulation split-up, first, the underlying complex biological mechanisms need to be understood and second, a data-based mathematical model describing population dynamics needs to be developed to bridge the gap from experimentally gained knowledge to optimization and control of industrial bioprocesses (Lencastre Fernandes et al., 2011). Here, the following work packages were designed in order to elucidate the role of environmental conditions in population dynamics:

- Choice of relevant industrial stress conditions
- Design of an experimental set-up to monitor population dynamics as a response to changing environmental conditions
- Accomplishment of three characteristic biological replicate cultivations for each stress condition
- Statistical assessment and analysis of flow cytometry and ‘whole transcriptome shotgun sequencing’ datasets
- Definition of a meaningful subpopulation separation variable
- Quantification of population dynamics as a response to alternating environmental conditions
- Formulation and implementation of a mathematical framework describing the observed population dynamics
- Combination of fermentation studies and a mathematical framework to decipher population dynamics during altering stress/stress-free cultivation conditions

1.2. *P. putida* KT2440 as a promising industrial production host

Selecting a suitable microorganism as a production chassis is a crucial step for the success of the production process (Lee et al., 2012). Ideally, a microbial cell factory is equipped with a variety of physiological and metabolic traits (Sauer et al., 2012): the platform strain should be robust, genetically stable and metabolically diverse (Foley et al., 2010). For economic reasons, the cell factory should also convert substrates into biomass and/or products efficiently and predictably, while showing simple culture media demands (Nikel et al., 2014a).

Albeit the evident need for a bacterial chassis uniting most of these desirable traits, only few hosts (often *E. coli* strains (Chen et al., 2013; Gopal et al., 2013; Mizoguchi et al., 2007)) are currently applied in industrial processes. Much of contemporary metabolic engineering approaches rely on the use of only a few bacterial hosts as working platforms (Danchin, 2012; Singh, 2014). However, organisms that are easiest to manipulate are often not the most suitable for specific industrial applications. Therefore, the implementation of novel biotechnological platform cells for industrial applications is currently the subject of intense research. In this thesis, *P. putida* KT2440 is explored as a promising alternative microbial cell factory.

Optimizing microbial cell chassis by streamlining the genome

The third research objective of this thesis quantifies the impact of streamlining the genome as a strategy to optimize the energetic demands and production capacity of the cell factory *P. putida* KT2440 (chapter 8). The concept of a suitable host for biotechnological applications is more or less reminiscent to that of a minimal microbial cell. All elements that are considered unnecessary for cellular functions other than replication and self-maintenance (e.g. prophages, flagellar genes, cell-to-cell communication devices) should be removed in order to increase the energetic efficiency and metabolic predictability.

Genomic editing tools (Martínez-García et al., 2011a; Silva-Rocha et al., 2013) have facilitated the construction of a number of reduced-genome variants derived from the wild-type strain *P. putida* KT2440. Recently, Martínez-García et al. (2014b) reported the construction of a flagella-less variant of *P. putida* KT2440 with attractive properties, such as an elevated NADPH/NADP⁺ redox ratio. Moreover, the physiological effects of erasing all viral DNA encoded in the *P. putida* KT2440 chromosome were explored in several mutants (Martínez-García et al., 2014a). While streamlining the bacterial genome gave rise to interesting physiological properties, the industrial worth of reduced-genome *P. putida* strains have not been systematically explored hitherto.

Here, the following work packages were designed to evaluate two multiple-deletion *P. putida* strains as cell factories for heterologous protein production and to compare their physiological characteristics to the parental strain:

- Transformation of the *P. putida* KT2440 derivative strains with a *gfp* expression plasmid that serves as a model for heterologous protein production
- Set-up and accomplishment of controlled batch and continuous cultivations in bioreactors (biological triplicates of all derivative strains, with and without carrying the production plasmid)
- Determination and comparison of industrially-relevant physiological parameters, such as specific growth rates, biomass and product yield coefficients and specific uptake- and production rates
- Determination and quantification of the maintenance demands of all derivative strains and assessment of their energetic household

CHAPTER 2

PSEUDOMONADS AS ORGANISMS OF INTEREST

The genus *Pseudomonas* comprises more than 200 species of Gram-negative gamma-proteobacteria with a respiratory rather than a fermentative metabolism (Palleroni, 1984; Timmis, 2002). *Pseudomonads* were first described as non-sporulating, polar flagellated rods by Prof. Migula of the Karlsruhe Institute in Germany at the end of the 19th century (Migula, 1894). The name is speculated to originate from the greek 'pseudes' (false) and 'monas' (single unit) (Palleroni, 1984).

Pseudomonads are found ubiquitously in the environment. Their extraordinary metabolic and genetic versatility –some species can metabolize more than 100 different sources of carbon and energy (Timmis, 2002)– allows them to adapt to different physicochemical and nutritional environments and populate highly diverse ecological habitats, ranging from natural environments over insects and plants to humans (Nikel et al., 2014a). Consequently, these bacteria are not only engaged in numerous important environmental activities, including degradation and recycling of organic compounds, but they also take part in food spoilage and parasitism and pathogenicity in plants, animals and humans (Timmis, 2002; Nikel et al., 2014a).

Apart from an exceptional metabolic diversity, *Pseudomonads* are known to be remarkable stress resistant, especially towards oxidative stresses (de Lorenzo, 2014). It was suggested that the Entner-Doudoroff pathway, which is exclusively used for sugar catabolism as a result of the absence of 6-phosphofructokinase activity, enables this high oxidative stress tolerance (Chavarría et al., 2013) by generating reducing equivalents at a high rate (Blank et al., 2008). Furthermore, *Pseudomonads* caught industrial attention because of their natural ability to produce bioactive compounds, such as antibiotics (Nikel et al., 2014a). Especially the non-pathogenic branch of *P. putida* species was taken into spotlight as promising platform strains (Nikel et al., 2014a; Poblete-Castro et al., 2012).

2.1. *P. putida* as industrial production host

P. putida strains are traditionally known as laboratory workhorses for the study of environmental bacteria because of their fast growth at simple nutrient demands and their complaisance towards genetic manipulation (Timmis, 2002; Nikel et al., 2014a). Inheriting the same metabolic versatility as all *pseudomonads*, *P. putida* strains are prominent for their resistance to antibiotics, disinfectants, detergents and even some heavy metals, while being capable to utilize aliphatic and aromatic hydrocarbons, which are toxic or hamper growth in most other microbial cell factories (Timmis, 2002; Schmid et al., 2001).

The starting point of the biotechnological career of *P. putida* was the discovery of its ability to degrade recalcitrant and inhibiting xenobiotics, such as toluene and xylenes, into central metabolites (Nakazawa et al., 1973). A key feature of *Pseudomonas*' metabolic diversity is their susceptibility towards transmissible plasmids and their rather relaxed-specificity gene expression system (Timmis, 2002). It does not come to a surprise, that only little time later more 'exotic' plasmid-encoded metabolic phenotypes, such as the capability to break down naphthalene (NAH7 plasmid in *P. putida* PpG7 (Dunn et al., 1973)), phenol (pPGH1 plasmid in *P. putida* H (Herrmann et al., 1987)) and 4-chloronitrobenzene (plasmid pZWL73 in *P. putida* ZWL73 (Zhen et al., 2006)) were discovered. Nonetheless, *P. putida* strains inherit also a wide range of chromosomally encoded catabolic degradation pathways and enzymes: In *P. putida* KT2440, alone >80 genes encoding oxido-reductases, which are needed for the metabolization of organic substrates, are present (dos Santos et al., 2004; Jiménez et al., 2002).

***Pseudomonas putida* KT2440** *P. putida* KT2440 is the plasmid-less derivative of the best-characterized toluene degrading *P. putida* mt-2 strain, that was first isolated in Japan (Bagdasarian et al., 1981; Nakazawa, 2002). In 1981 strain KT2440 was certified as the first Host-Vector Biosafety system for gene cloning in Gram-negative soil bacteria by the Recombinant DNA Advisory Committee of the U.S. National Institute of Health. Lacking any pathogenesis determinants, it was also Generally Recognized as Safe (GRAS) by the U.S. Food and Drug Administration.

Being considered as one of the safest and most secure hosts for foreign gene cloning, *P. putida* KT2440 emerged as the workhorse of soil bacteria/*P. putida* research. The advance of the genomic sequence (Nelson et al., 2002) revealed various genetic determinants playing a role in biocatalysis and industrially relevant enzymes, such as the production of epoxides, substituted catechols, enantiopure alcohols, and heterocyclic compounds (Wackett, 2003). In combination with genome-wide pathway modeling (Puchařka et al., 2008; Nogales et al., 2008; Sohn et al., 2010) the way was paved for advanced engineering strategies and systems biology approaches (Reva et al., 2006). System-wide analysis has been shown to be a powerful tool to provide a solid knowledge-base on

metabolic and regulatory features. But, so far, the majority of available related studies focused on degradation processes (Puchałka et al., 2008).

Recently, research collaborations started to focus on biotechnological applications using metabolic engineering or a combination of multi-omics studies and systems-wide metabolic modeling towards increasing and promoting *P. putida*'s performance as a cell factory of choice for white biotechnology (Verhoef et al., 2010). Unfavorable traits for industrial applications, such as the lack of a fermentative metabolism and a rather high abundance of mobile genetic elements have been tackled recently (Nikel et al., 2012; Martínez-García et al., 2014a). To date, strains of *P. putida* have been engineered to produce biobased polymers and a wide range of chemicals such as phenol and *p*-hydroxybenzoate (Wierckx et al., 2005; Meijnen et al., 2011). Furthermore, enzymes from *P. putida* found industrial application in a variety of biocatalytic processes (Schmid et al., 2001). However, most *Pseudomonas*-based applications are still in their infancy and industrial key processes are still dominated by *Bacillus*, *Corynebacterium glutamicum* and *Escherichia coli* (Puchałka et al., 2008).

CHAPTER 3

CHARACTERIZATION OF MICROBIAL POPULATIONS

Traditionally, isogenic microbial cultures are considered to be uniform: Only little morphological and physiological diversity is assumed within individual cells of one population. However, recent research discovered that individuals within a population are by no means identical (Müller et al., 2010; Avery, 2006). Regarding industrial production cultures, a heterogeneous population might contain poorly producing subpopulations, which will negatively impact the overall productivity. Until now, the underlying mechanisms that are suspected to give rise to population split-up are neither completely understood nor included when developing new bioprocess control strategies or optimizing existing fermentation strategies (Lencastre Fernandes et al., 2011). Optimization approaches are neglecting differences in phenotypes of single cells, leveling cell properties to averages. For example, the specific productivity that is observed in a biotechnological process could result from different population compositions. On one hand, the population could be uniform, containing individuals that deviate from the mean. But on the other hand, the population could also be composed of two or more subpopulations characterized by a specific productivity, each one different from the mean (Figure 3.1). Therefore, valuable information about population dynamics is camouflaged and optimization approaches might be misled due to false assumptions about the population state.

This chapter summarizes (i) the emergence and suggested origins of population heterogeneity and (ii) experimental methods to assess individual cell behavior in microbial populations.

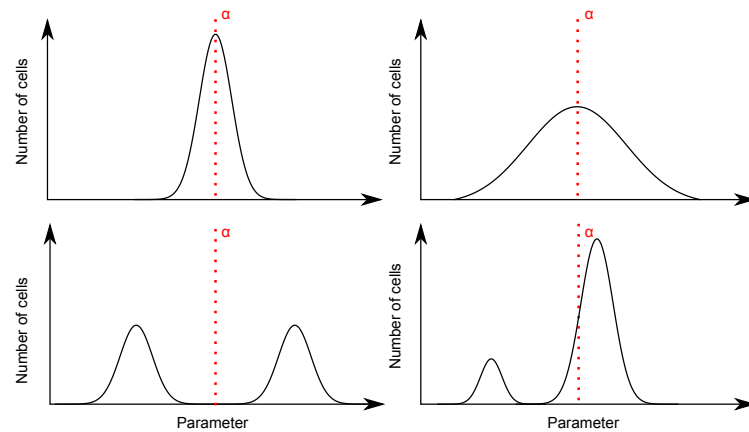


Figure 3.1.: Schematic description of how average measurements mask real population states. Typically, an average value α of a quantifiable parameter describing the population state, e.g. the specific productivity, is measured when monitoring a microbial population. Notably, the measured average value α can result from differently composed cell populations, either from uniform populations, differing in deviation from the mean (a and b), but also from differently distributed subpopulations (c and d). Therefore, average values of populations mask real parameter distributions of single cells and might cause misleading optimization strategies of microbial cultivations (adapted from Dhar and McKinney (2007)).

3.1. The origin of population heterogeneity in clonal bacterial populations

Heterogeneity of clonal microbial cultures may result from several distinct sources, including manifold biological, but also several technical factors (Müller et al., 2010) (Figure 3.2). Differences in growth and cell cycle states, gene expression noise and asymmetric cell division are considered as origins of heterogeneous populations. Further biological factors suggested to be implied in the emergence of subpopulations are gene mutations or loss, variability in plasmid copy numbers and epigenetic modifications (Müller et al., 2010; Fritzsche et al., 2012; Jahn et al., 2012). External factors, such as fluctuating environmental conditions due to deficient mixing in large-scale industrial reactors play an additional role in the advent of population heterogeneity (Schweder et al., 1999).

The impact of many of these factors are difficult to address at the single cell level and their direct quantitative influence on population heterogeneity remains rather unclear (Avery, 2006). Here, we focus on the cell cycle as a biological factor and industrially-relevant stress environments as an external technical factor promoting the advent of population heterogeneity.

The cell cycle as origin of population heterogeneity

In the context of this thesis, the impact of the cell cycle on population heterogeneity will be quantified. Cell cycle decisions and variations in the duration of different cell cycle phases are

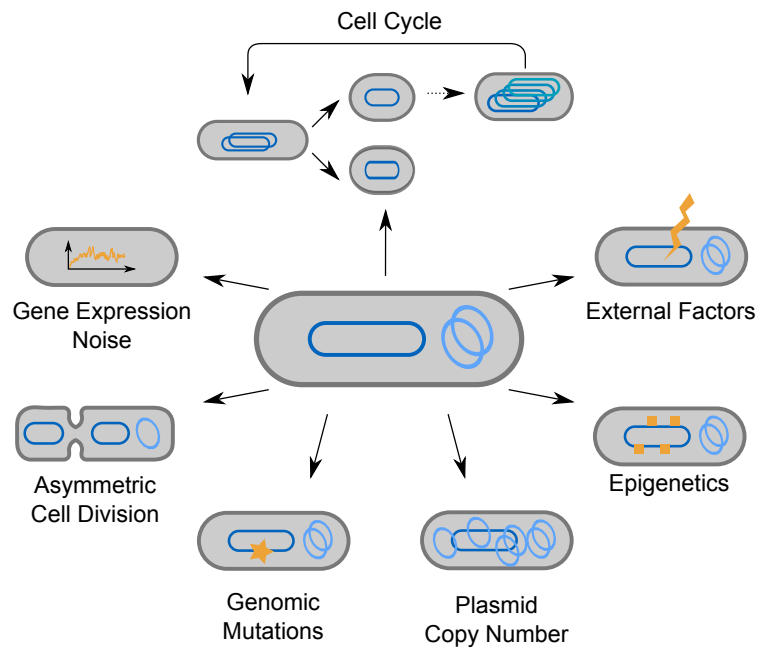


Figure 3.2.: Overview of origins of population heterogeneity. Many biological and external factors are considered to cause population heterogeneity (adapted from Jahn et al. 2012).

considered to play a key role in the advent of population heterogeneity (Müller et al., 2010).

The bacterial cell cycle consists of 4 distinct phases (described for *Escherichia coli*, Figure 3.3): The first phase, the B phase, is defined as the time between division and start of replication. It is followed by the replication phase (C phase), the pre-D phase –an interphase between the C and D phase– and the division phase (D phase) (Cooper, 1991; Müller et al., 2003). The durations of the replication and cell division phases (C and D phases) were found to be relatively independent of the growth conditions and are therefore assumed to be constant (Cooper et al., 1968). Contrary, the interphases of the cell cycle (B and pre-D phases) are subject to much variation (Müller, 2007). In *E. coli* the duration of the B period is coupled to a constant critical bacterial cell mass (Donachie, 1968). This critical cell mass is already existing or rapidly reached by the cell under nutrient-rich conditions, while more time is needed in nutrient poor media. The pre-D phase specifies the bacterial disability to divide after finishing replication, obviously ‘waiting’ for permissive growth conditions (Müller, 2007). Consequently, the pre-D phase disappears under optimal growth conditions similar to the B phase. Under nutrient-rich conditions some bacterial species can accelerate proliferation and decrease their generation time below the sum of their C and D phases: new rounds of DNA replication is initiated before a previous round has been completed (Cooper et al., 1968).

In recent years it was discussed that metabolic activity might differ in dependency of specific cell cycle phases. Studies of *Methylobacterium rhodesianum* showed that products like polyhydroxyalkanoates (PHAs) only accumulate when cells harbor a specific amount of DNA equivalents,

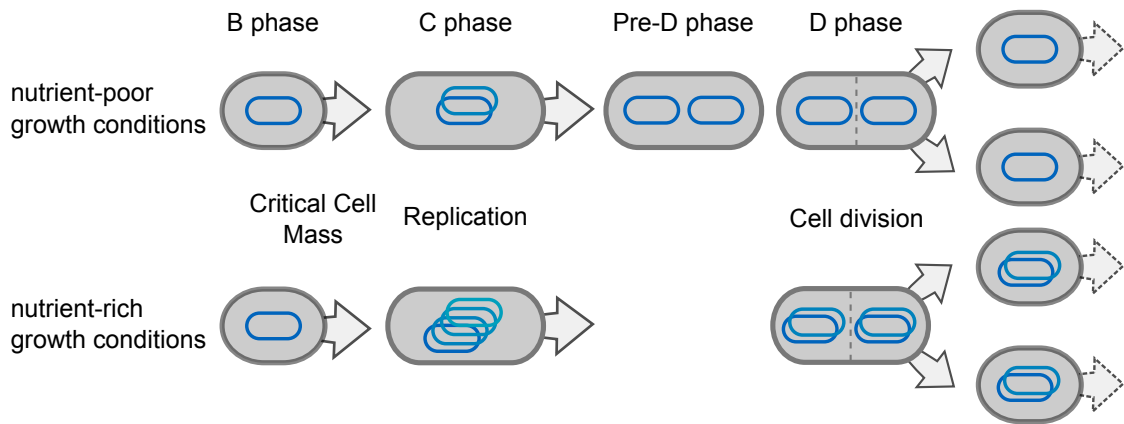


Figure 3.3.: Schematic overview of the bacterial cell cycle. The bacterial cell cycle can be divided into B, C, pre-D and D phases. Under unlimited growth conditions, some bacterial species are capable of accelerating proliferation by uncoupling DNA synthesis from division. As a result, a new round of DNA replication is initiated before the completion of the previous round (Cooper, 1991; Müller et al., 2010).

reflecting a specific cell cycle phase (Ackermann et al., 1995). This phenomenon was found to occur at off-cell-cycling stages and started the discussion, if e.g. product synthesis might only take place during the stochastic B- and pre-D phases, when the cell is neither replicating nor dividing (Bley, 1990; Müller et al., 2010). In chapter 6 we investigate the dependency of the cell's protein inventory on cell cycle stages and how growth rates may influence both, protein composition and cell cycling.

Heterogeneity enforced by environmental conditions

Additional to the biological side, technical settings can cause population heterogeneity. Growth conditions in laboratory-scale bioreactor cultivations are usually designed to be ideal: well-mixed and homogeneous. However, during large-scale industrial fermentations, which aim for high biomass concentrations and/or high product yields, cells are often exposed to less ideal conditions. Mixing times can rise to the order of minutes in viscous fermentation broths. Cells circulating through a bioreactor are therefore exposed to different local environments, e.g. zones of varying dissolved oxygen availability (Gelves et al., 2014). Therefore, each individual cell 'sees' different environments during its generation time in the bioreactor. Continuously changing micro-environments may cause repeated cycles of induction and relaxation of stress responses, adaptational processes or metabolic adjustments. Physiological properties of the microbial population may alter during the production process, resulting in subpopulations with reduced biomass yields and productivities (Schweder et al., 1999; Lara et al., 2006; Enfors et al., 2001; Lencastre Fernandes et al., 2011; Carlquist et al., 2012). In chapter 7 we investigate and quantify the impact of challenging industrially-relevant environmental conditions on population dynamics.

3.2. Cultivation strategies and experimental methods for deciphering population dynamics

Deciphering population dynamics and highlighting differences in cell behavior on the single cell level is dependent on a reliable and carefully designed experimental and analytical set-up. The following sections give an overview about cultivation strategies and differences in experimental and analytical methods for the characterization of average populations and microbial heterogeneity.

3.2.1. The chemostat as a model system to decipher population dynamics

The key for deciphering and quantifying the impact of one driver of population dynamics is an experimental set-up, which allows to specifically change one single parameter, keeping all other cultivation parameters constant. Furthermore, in order to collect reproducible and reliable datasets, cells have to be grown in a defined, constant and controllable set of physico-chemical conditions (Hoskisson et al., 2005).

Bioreactor cultivations provide controlled environmental conditions. Several operating modes can be realized: discontinuous ‘batch’ and ‘fed-batch’ or ‘continuous’ cultivations. Discontinuous cultivation systems result in dynamic physico-chemical conditions. Datasets are therefore often complex and difficult or impossible to interpret when trying to dissect the influence of one specific parameter (Hoskisson et al., 2005). For example, the typical microbial growth curve, consisting of different growth rates in lag, exponential and stationary phases, is not an inherent property of the organism but a result of its interaction with the constantly changing physico-chemical environment in which it is growing in batch cultivation (Tempest, 1970). Contrary, the chemostat offers the great advantage of uncoupling growth rate from the transient conditions encountered in batch culture by providing constant cultivation conditions (Bull, 2010).

Introduced simultaneously by Novick and Szilard (1950) and Monod (1950), the chemostat is the most commonly used experimental approach for investigations of physiology in steady-state cultures (Bull, 2010).

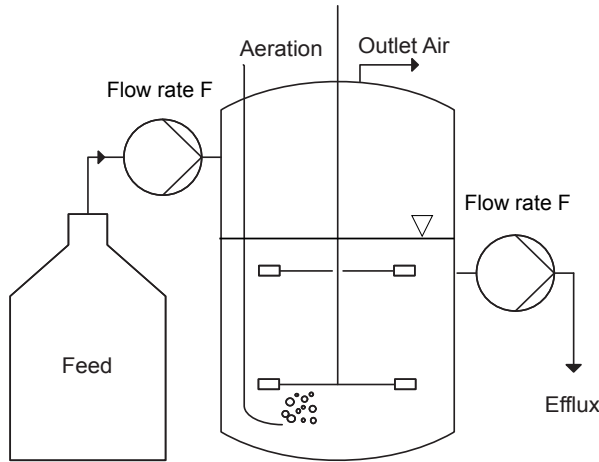


Figure 3.4.: Schematic chemostat set-up. The influx of feed medium and the efflux of cultivation broth is constantly balanced ($F_{in} = F_{out}$, in Lh^{-1}), keeping the reaction volume constant.

In a chemostat, the feed of sterile medium from a reservoir is balanced by the efflux of spent medium, living cells, cell debris and excreted products (Figure 3.4). The flow rate of medium F (in Lh^{-1}) into the vessel is related to its cultivation volume V (in L) and defines the dilution rate D (in h^{-1}):

$$D = \frac{F}{V} \quad (3.1)$$

The chemostat device allows growth to occur at an equilibrium, called steady state, where growth of new cells (biomass X) is being balanced by those washed out.

$$\frac{dX}{dt} = \mu X - DX \stackrel{!}{=} 0 \quad (3.2)$$

This means that the growth of new biomass is equal to the rate at which the culture is being diluted. Hence, establishing steady-state conditions $\frac{dX}{dt} = 0$ result at equal growth rate μ and dilution rate D .

The possibility to manipulate the specific growth rate of the organism externally by setting a specific dilution rate is the key feature of the chemostat. It makes it a versatile tool to individually change one culture parameter, while all other relevant physical and chemical culture parameters are kept constant (composition of synthetic medium, pH, temperature, aeration, etc.) (Hoskisson et al., 2005). The use of chemostat cultures in the fields of basic physiology and biochemistry led to milestones in our understanding of the basis of microbial processes ((Monod, 1949; Herbert et al., 1956; Pirt, 1965) and many more). However, with the advent of molecular biology research, employing chemostats as the cultivation mode of choice faded into the background. Only during the last 10 years, focussing on global and systems level investigations of the organism, continuous cultivation at steady state conditions made a comeback (Ferenci, 2006).

In conclusion, chemostat studies allow the targeted investigation of the impact of one specific parameter and advantages in environmental control (Hoskisson et al., 2005). For this study, chemostat cultivations offer the ideal system to investigate origins of population heterogeneity. The experimental set-up provides an environment in which cell division is continuous but population size is held constant. Under steady state conditions, it is possible to analyze and compare the composition of a population quantitatively, enabling the deciphering of the impact of e.g. cell cycle and differing environmental conditions on population dynamics, one at a time. Additionally, the

carefully controlled and defined physiological conditions obtainable in chemostat cultures allow the acquisition of reproducible and reliable biological samples (Wu, 2004).

3.2.2. Experimental and analytical methods for describing cell populations and distributions of cell properties

Understanding the functional structure and dynamics of a cell requires the investigation of its biology at a systems level (Kitano, 2002). One of the essential properties of a cellular system is its robustness (Csete et al., 2002). Mechanisms and principles that determine the robustness of the cell –for instance adaptation during exposure to challenging environments– need to be investigated for a systems understanding to ultimately guide the design of robust microbial cell factories. The basis of systems-level analysis is the acquisition of a comprehensive set of quantitative data (Kitano, 2002).

‘Omics’ approaches Omics studies aim at the precise quantification of pools of biological molecules that translate into the structure, function and dynamics of the cell. The availability of genome sequences paired with precise high-throughput measurements of e.g. proteins (proteomics), RNA (transcriptomics) or metabolites (metabolomics), enables the collection of comprehensive data sets on the performance of the cell and gaining information on the underlying mechanisms (Kitano, 2002). Among the different levels of ‘omics’ approaches, transcriptome analysis is the detection and quantification of an ideally complete set of transcripts in a cell at a specific time point and environmental condition. The transcriptome is a quantitative measure of the global expression level of mRNA molecules and is therefore indicative of gene activity (Adams, 2008). The comparison of differentially expressed genes at various environmental conditions can give insights into underlying mechanisms of stress response and adaptation.

Another powerful ‘omics’ approach is the quantification of the protein content of the cell. Proteome analysis gives a ‘snapshot’ of the total cellular protein abundance at a given time point and environmental condition. Analyzing the global protein content is a tool to measure cellular functionality (Tyers et al., 2003). Protein patterns reflect cell decisions and the analysis of the proteomic inventory of a cell can unravel metabolic characteristics and responses to e.g. challenging environments (Jahn et al., 2012).

However, these classic ‘omics’ approaches only give insight into the mRNA expression levels or protein abundance of the average population (Müller et al., 2010). Heterogeneity within a clonal population is not considered and cannot be quantified. In order to dissect differences on a single cell level, two options are available: (i) limiting the analytical space to a dimension of a single cell in a lab-on-a-chip device (Fritzsche et al., 2012) or (ii) analyzing specific parameters of cells in a population one by one in high-throughput (Schmid et al., 2010).

Flow cytometry Cytometry deals with the measurement of optical properties of single cells. Flow cytometry examines particles, such as cells, in suspension instead of statically under the microscope (Müller et al., 2010). The method, originally developed for the detection of aerosolic biological weapons (Gucker et al., 1949), was further developed and applied the first time in the 1960's for biological studies (Kamentsky et al., 1965; Dittrich et al., 1969).

Today, flow cytometry and flow cytometry-based cell sorting techniques have become indispensable in cell and developmental biology. The methods allow us to obtain information about phenotypic diversity of individual cells within a population. In combination with a sorting device, individual cells can not only be characterized but additionally sorted according to specific parameters (Müller et al., 2010).

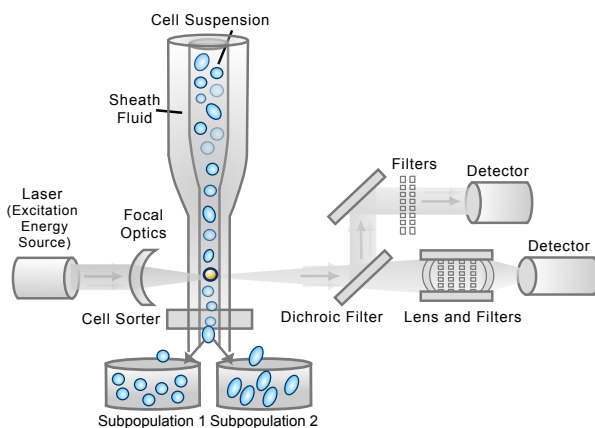


Figure 3.5.: Schematic flow cytometer set-up.

In the flow cytometer, cells are transported in a laminar fluid stream to a laser beam –one cell at a time at high speed (Figure 3.5). Cell properties are interrogated in the flow cell at the laser intercept: one or multiple laser beams illuminate the cells in order to measure their light scattering properties and/or to excite fluorescent molecules. Light scattering data contains information about the relative cell size (forward scatter) and cell granularity or internal complexity (side scatter) (Müller et al., 2003).

Additionally, suitable lasers can excite and give quantitative information on fluorescent molecules either naturally occurring inside the cell or specifically employed to tag or stain cellular components. A variety of dyes have been employed to study cellular parameters, such as intracellular pH, membrane potential or the levels of cellular components, e.g. DNA (Müller et al., 2003). A detailed overview of available dyes is beyond the scope of this section and is given elsewhere (Nebe-von Caron et al., 2000; Shapiro, 2000; Lencastre Fernandes et al., 2011). The ability of studying diverse parameters in a high-throughput manner at the same time enables statistically robust analysis of parameter distributions throughout the cell population and forms the core of the method (Lencastre Fernandes et al., 2011).

A flow cytometer can be also equipped with a cell-sorting device for Fluorescence-Activated Cell Sorting (FACS). Sorting decisions can be based on one or multiple cellular properties. The sorted cells might, if the staining method allows, be cultured for further molecular or functional assays or enrichment (Tracy et al., 2010).

Flow cytometry was proven to be a suitable tool to quantify the distribution of characteristics of individual cells within a population (Skarstad et al., 1983; Srienc, 1999; Shapiro, 2000; Müller

et al., 2003; Brehm-Stecher et al., 2004). Consequently, flow cytometry data sets were used in this thesis to derive experimental data on population dynamics.

However, flow cytometry methods require an intact cell wall structure of the organism. Therefore, these approaches do not allow access to holistic quantitative data on the cell's 'interior' in contrast to 'omics' approaches (Müller et al., 2010). The contradiction between the unapproachability of the e.g. global protein abundance within a cell by flow cytometry and missing information on heterogeneity by 'omics' approaches can be solved by a careful combination of these techniques: First, cells are sorted into subpopulations according to a distinguishing parameter that was characterized via flow cytometry. Second, the protein inventory of the different subpopulations is investigated, using classical proteomic methods (Jehmlich et al., 2010).

Here, we combine carefully designed chemostat cultivations with flow cytometry, cell sorting and proteomics to investigate the impact of the cell cycle on the onset of population heterogeneity.

CHAPTER 4

MODELING MICROBIAL POPULATIONS

During the last decades, a variety of mechanistic models differing in the degree of complexity have been applied to describe cell populations and cultivations quantitatively. Fredrickson et al. (1967) formulated a systematic framework for viewing cell populations, dividing mathematical descriptions into ‘segregated’ or ‘unsegregated’ and ‘structured’ or ‘unstructured’ (Figure 4.1).

Briefly, ‘segregated’ models describe individual differences of cells within a population whereas ‘unsegregated’ models assume an uniform average cell population. Furthermore, ‘structured’ models account for differences in chemical composition inside the cells, whereas ‘unstructured’ models describe cells as uniformly composed ‘black boxes’.

Unsegregated models Most commonly, *unsegregated* and *unstructured* models are applied for quantifying biological cultivations (Nielsen et al., 1992). This type of model is the most idealized view on biomass: ‘Average’ cells are described, considering only external input and output variables, such as substrate uptake and production, while intracellular kinetics are neglected.

Unsegregated structured models resolve biomass to a higher degree. *Structure* can be defined either in a physical sense, such as division into organelles, shape or size of a cell, or in a biochemical sense, where biomass is subdivided into its intracellular biochemical components (Gernaey et al., 2010). These kind of models have been successfully applied to describe intracellular metabolism or filamentous growth (Gombert et al., 2000; Stelling, 2004).

Segregated models consider population heterogeneity by accounting for property distributions of single cells (Fredrickson et al., 1970). *Segregated unstructured* models consider the co-existence

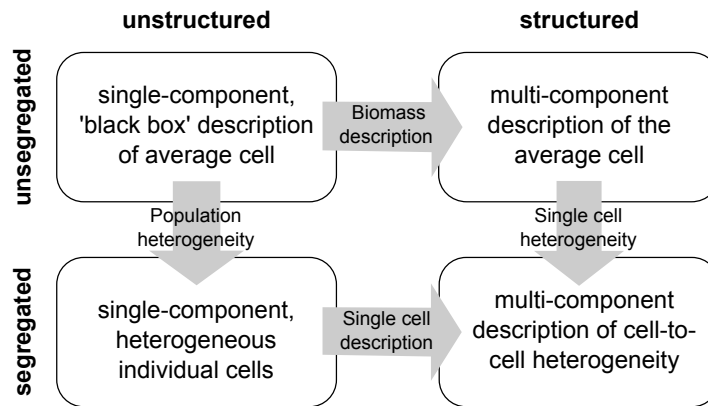


Figure 4.1.: Classification of mathematical models used to describe microbial cultivations. Mathematical frameworks differ in the degree of complexity. Segregated models account for heterogeneity instead of assuming averaged cell behavior, while structured models consider more than one cellular component in their population/cell description (Adapted from Lencastre Fernandes et al. (2011)).

of subpopulations without describing details of their e.g. intracellular composition (Lencastre Fernandes et al., 2011). *Structure* can be introduced by accounting for at least one distinction variable, such as cell mass or age (Ramkrishna, 2000). The degree of complexity can be deepened to certain extends in multivariate models. In case of chemically structured models, multiple metabolites are accounted for as internal state variables (Bailey, 1998). Segregated and structured models are therefore able to describe subpopulations that differ in e.g. the phase of the cell cycle (Fredrickson, 2003) or production and non-production states (Mantzaris et al., 2002).

Various mathematical formulations have been proposed to describe distributed properties, among them ordinary or delay differential equations and population balance models (PBMs) (Bley, 2011).

The most common types of PBMs are population balance equation models (PBE) and cell ensemble models (Sidoli et al., 2004; Henson, 2003). The two modeling strategies differ in depth of intracellular description and number of cells included in the model: While PBEs allow modeling of large populations, only a small number of variables –often a single variable, such as cell age (Sherer et al., 2008) or mass (Hatzis et al., 2006)– can be used to characterize the intracellular state of the cell. In contrast, cell ensemble models are limited in the cell number of the population, but are constructed from more complex single cell models, allowing detailed description of the intracellular state (Henson, 2003).

Mathematically, a PBE includes a dynamic cell balance, which is formulated as a nonlinear partial differential equation (Fredrickson et al., 1967). A distribution function $f(\mathbf{y}, t)$ describes the distribution of cells characterized by the internal state vector \mathbf{y} at time t . The quantity of cells in the population $n(t)$ per property space $V_{\mathbf{y}}$ can therefore be calculated as follows:

$$n(t) = \int_{V_{\mathbf{y}}} f(\mathbf{y}, t) d\mathbf{y} \quad (4.1)$$

In most cases, a single limiting substrate is considered and its mass balance allows the determination of n at steady state, considering the volumetric substrate consumption rate $r_s(\mathbf{y}, s)$ as being independent of the physiological state (Villadsen et al., 2011):

$$\frac{ds}{dt} = D(s_0 - s) - \int_{V_y} r_s(\mathbf{y}, s) f(\mathbf{y}, t) d\mathbf{y} \quad (4.2)$$

$$n = \frac{D(s_0 - s)}{r_s(s)} \quad (4.3)$$

A dynamic population balance equation for the cell distribution can be set up as follows (Villadsen et al., 2011):

$$\underbrace{\frac{\partial f(\mathbf{y}, t)}{\partial t}}_{\text{accumulation}} + \underbrace{\nabla_{\mathbf{y}}[\mathbf{r}(\mathbf{y}, t) f(\mathbf{y}, t)]}_{\text{single cell growth}} = 2 \underbrace{\int_{V_y} b(\mathbf{y}^*, t) p(\mathbf{y}, \mathbf{y}^*, t) f(\mathbf{y}^*, t) d\mathbf{y}^*}_{\text{birth}} - \underbrace{b(\mathbf{y}, t) f(\mathbf{y}, t)}_{\text{division}} - \underbrace{D f(\mathbf{y}, t)}_{\text{dilution}} \quad (4.4)$$

The balance is coupled to mathematical formulations of (i) the single cell growth rate $r(\mathbf{y}, t)$, (ii) the breakage function $b(\mathbf{y}, t)$ and (iii) the partitioning function $p(\mathbf{y}, \mathbf{y}^*, t)$ (Srienc, 1999). These three functions define the dynamics of a cell type, where $r(\mathbf{y}, t)$ describes the rate of accumulation of a property within a cell, $b(\mathbf{y}, t)$ represents the probability of cell division at the physiological state \mathbf{y} and $p(\mathbf{y}, \mathbf{y}^*, t)$ specifies the probability of a cell at state \mathbf{y}^* to divide into two daughter cells with the physiological states \mathbf{y} and $\mathbf{y}^* - \mathbf{y}$ (Stamatakis, 2010). D is the dilution rate of the reactor. Here, we assume that no living cells enter the reactor with the feed medium and that the reactor is a homogeneous environment.

Getting closer to reality The models introduced in the previous paragraphs assume homogeneous growth environments of the cell population. In case of large-scale industrial cultivations, this simplification cannot be sustained. Spatial heterogeneity can be introduced by coupling metabolic network modeling or even population balance modeling to computational fluid dynamics. Thereby, it is possible to reflect the interplay of intracellular and environmental variations, getting one step closer to a realistic description of a bioreactor (Lapin et al., 2004).

The application of population balance modeling has increased exponentially during the last two decades (Ramkrishna et al., 2014) and biotechnological relevant examples have been comprehensively reviewed by Lencastre Fernandes et al. (2011). However, to date we are still far away from using PBM approaches as a standard tool for bioprocess optimization due to several challenges ahead: Besides formulation and solution of models with multidimensional state vectors, where increased computing power and development of efficient methods for discretization and numerical solution are needed (Mantzaris et al., 2001a; Mantzaris et al., 2001b; Mantzaris et al., 2001c), the biggest difficulty lies in the definition of the physiological state functions described above (Henson,

2003). Single cell analysis, such as flow cytometry, is needed to determine the single cell growth rate and the division and partitioning function systematically in order to minimize the number of assumptions introduced into the model (Lencastre Fernandes et al., 2011).

Selection of the mathematical model Clearly, a structured and segregated model accounting for spatial heterogeneity would offer the most realistic representation of a cell population in a bioreactor. However, there are important trade-offs to be made in terms of formulation time, model complexity, and solution time when setting up the mathematical framework (Sidoli et al., 2004).

The choice of a suitable mathematical framework out of the “zoo of mathematical models” available (Bailey, 1998) is dependent on the purpose of the model. Considering the question “What is the specific problem the model is supposed to solve?” prevents ending up with a mathematical description reasonably reflecting experimental data, but with no informational aspects or biological conclusion (Casti, 1997). Furthermore, the governing principle is to keep the model as simple as possible, while including all essential information to explain the observed phenomena (Villadsen et al., 2011).

In this thesis, two different kind of models were applied for two different purposes. On one hand, the question of how environmental factors influence population heterogeneity was addressed. Here, a ‘segregated’ view on the cell population is needed to account for different subpopulations within one cell population. Flow cytometry data was systematically used in a mechanistic model to describe the relationship between environmental stress conditions and population heterogeneity. The major advances made in single cell analysis during the recent years have so far not been translated to the same extend into advances in modeling of population heterogeneity. Here, we take one step in the direction of closing this gap by deciphering one mechanism of the environment as a driver of cell heterogeneity. The mechanistic model together with the experimental flow cytometry data gathered can be seen as a basis for a set up of a corresponding population balance model, giving valuable information on the calculation of the physiological functions as well as the state vector. The model used to calculate the influence of environmental conditions on population heterogeneity and its computational implementation is described in section 7.3.1.

On the other hand, in chapter 8, industrially relevant physiological parameters of genome-reduced *P. putida* derivative strains needed to be assessed in order to compare their growth and production performance to the wild type strain KT2440. Here, unstructured models, as simple as mass balances applied to microbial growth and product formation, provide a simple but useful description of growth kinetics and production capacity (Roels, 1980; Brass et al., 1997). The mathematical framework used is explained in detail in section 5.6.

CHAPTER 5

MATERIAL AND METHODS

This chapter gives an overview about materials and methods that have been used throughout this thesis. Parts of this chapter have been submitted partially or in detail for publication. Cross-references to the manuscripts (Appendices A - C) are provided.

5.1. Bacterial strains, media and cultivation systems

In this thesis, different *P. putida* strains were used for specific research aims as listed:

- The wild-type strain *P. putida* KT2440 (ATCC47054, Bagdasarian et al., 1981), acquired from the ‘Leibniz Institute DSMZ’ (German Collection of Microorganisms and Cell Cultures), was used as a model strain for the studies of population heterogeneity (chapters 6 and 7).
- A summary of the *P. putida* derivative strains and plasmids used in chapter 8 (‘Optimizing microbial cell chassis by streamlining the genome’) is provided in Table C.1 in the appendix. Here, the laboratory wild-type strain *P. putida* KT2440 that served as a basis for the derivative strains was kindly provided by Víctor de Lorenzo (CNB-CSIC, Madrid) and was used as a reference in this specific study for consistency reasons.

Detailed information about pre- and main culture media compositions, as well as about seed train procedures and process conditions can be found as follows:

- All pre- and main cultivations were carried out using minimal medium M12, as described in section A.2 ‘Bacterial strains and cultivation conditions’.

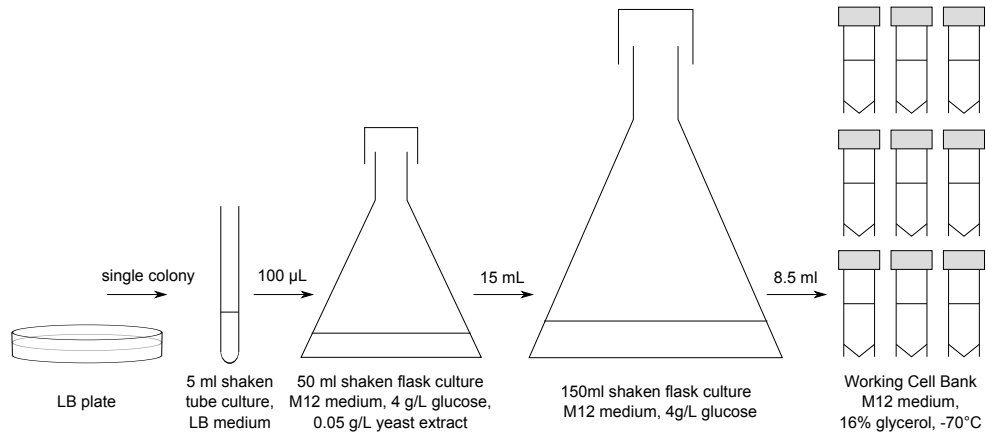


Figure 5.1.: Schematic working cell bank (WCB) procedure For every strain that was used in this thesis, a WCB was established, derived from a single colony on a LB plate. Cells were grown, stepwise reducing the complex medium content until grown in M12 minimal medium. Cells were harvested in mid-exponential phase and stored as WCB at -70°C in a 16 % (v/v) glycerol stock

- To minimize population heterogeneity at the starting point of the pre-/main cultivation, bioreactor batch and continuous cultivations were inoculated with a cryogenic working cell bank, derived from a single colony on a LB plate, afterwards grown and harvested from exponential phase cultures, stepwise reducing the complex medium content, until grown in M12 minimal medium and stored as a working cryo-culture bank at -70°C in a 16 % (v/v) glycerol stock (Figure 5.1).
- Batch and continuous cultivations were carried out in a 3.7 L scale bench-top reactor (KLF, Bioengineering, Switzerland) at a working volume of 1.5 L. A schematic set-up of the bench-top reactor as a chemostat can be found in Figure 5.2. In case of batch cultivations the reactor set-up was used without feed and harvest installations. Detailed information about process conditions can be found in the appendix in section C.2 ‘Bioreactor cultures’.

5.2. Nucleic acid manipulation and plasmid construction

DNA manipulations used for the construction of the *gfp* expressing *P. putida* recombinants in chapter 8 followed well established protocols (Green et al., 2012). *P. putida* derivative strains were transformed with the *gfp* expression plasmid pS234G by electroporation (Choi et al., 2006). Detailed information about the construction of the expression plasmid pS234G can be found in the appendix in section C.2 ‘Nucleic acid manipulation, plasmid construction, and plasmid stability assay’ and Table C.1.

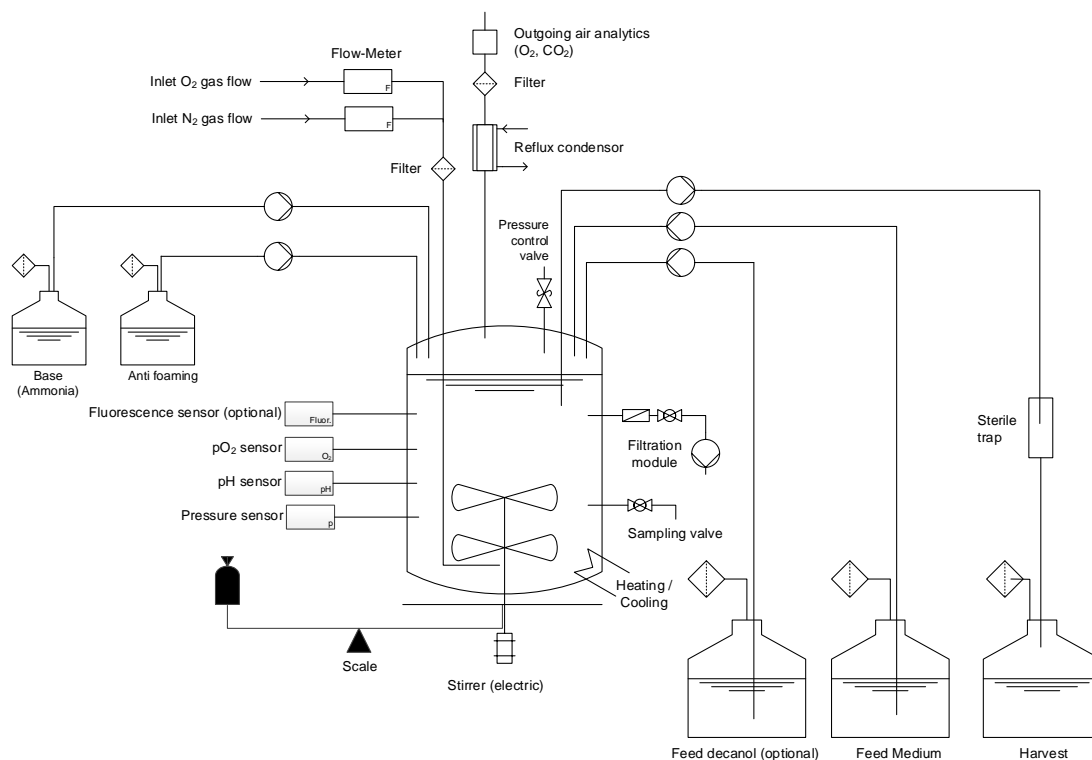


Figure 5.2.: Schematic 3.7 L bench-top reactor set-up. A 3.7 L bench-top reactor was run at a working volume of 1.5 L as batch and continuous cultivation set-up. Continuous cultivation was controlled gravimetrically. Medium was fed continuously into the bioreactor at a selected flow rate and culture broth was harvested repeatedly after a weight gain of 10 g was monitored. An additional feed of decanol was used in the solvent stress exposure studies (chapter 7). Batch cultivations for the physiological characterization of the *P. putida* recombinants (chapter 8) were carried out in the same reactor system, but without any feed or harvesting devices. Optional settings, e.g. an online fluorescence sensor, were used during cultivation of GFP expressing *P. putida* strains (chapter 8). Specifications of sensors and other parts used can be found in the appendix in section C.2 ‘Bioreactor cultures’.

5.3. Analytical methods

The cultivations carried out during this thesis were monitored in detail. Methods were applied consistently throughout all cultivations as follows:

- Analytical methods applied in all cultivations included the determination of biomass cell dry weight (CDW), optical density (OD₆₀₀) and concentrations of organic acids and nucleotides via high pressure liquid chromatography (HPLC). Detailed procedures are explained in the appendix in section C.2 ‘General procedures’ and ‘Analytical procedures’.
- The two-phase decanol/M12 medium cultivations carried out in chapter 7 required adjustments of biomass determination. Emulsion forming prohibited a reliable determination of the biomass by gravimetric or spectrophotometric methods. Therefore, biomass was determined using a moisture analyzer MB35 by Ohaus Europe GmbH (Switzerland). 5 mL of biomass suspension was heated to 120 °C for 120 min to remove all evaporable components. A filtrate sample was treated accordingly and the biomass was determined as weight difference between the treated biomass and filtrate samples. Additionally, the cell concentration was determined via cell counting in a counting chamber under a microscope at 400x magnification after adequate dilution (1:10 - 1:100). Every sample was counted three times and the arithmetic mean and standard deviation were calculated. To evaluate the reliability and accuracy of the cell counting method, OD₆₀₀ measurements and cell counts were correlated under standard conditions (no decanol addition).
- GFP fluorescence was quantified by spectrofluorimetry for the characterization of the *P. putida* KT2440 derivative strains in chapter 8. Fluorescence in samples of biosuspension and filtrate was quantified at 485 nm (excitation) and 535 nm (emission) in a fluorescence microplate analyzer (Synergy 2, BioTek Instruments, Inc., VT, USA) according to the protocol described in detail in the appendix in section C.2 ‘GFP quantification’.

5.4. Flow cytometry analysis, cell sorting and subpopulation-proteomics

Samples for flow cytometry analysis were taken throughout all cultivations. The flow cytometry measurements were carried out at the cooperation partner UfZ Leipzig using a MoFlo cell sorter (Beckman-Coulter, USA), as described before (Jahn et al., 2013). Forward scatter, side scatter and DAPI fluorescence were detected according to specifications given in the appendix in section A.2 ‘Sample preparation and staining for flow cytometry’ and ‘Flow cytometry and cell sorting’.

Flow cytometry data analysis The resulting data sets were analyzed as a part of this thesis using the statistical software R Bioconductor (www.bioconductor.org). The ‘gating’ process, which excludes technical noise, cell debris and agglomerated cells from the data set and selects different subpopulations was carried out using the packages flowCore (version v.1.11.20 (Ellis et al., 2014)) and flowViz (version v.0.2.1 (Ellis et al., 2013)).

Fluorescence activated cell sorting and subpopulation proteome analysis Cell sorting and proteome analysis were carried out for the investigation of the cell cycle as a driver of population heterogeneity (chapter 6). Cell sorting and identification of proteins by LC-MS-MS were carried out at the UfZ Leipzig as explained in the appendix in section A.2 ‘Flow cytometry and Cell sorting’ and ‘Identification of proteins by LC-MS-MS’.

The software MaxQuant (v1.2.2.5, (Cox et al., 2008)) was used to analyze mass spectra for protein identification and label-free quantification (LFQ) with the genome database of *P. putida* KT2440 (according to Jahn et al. 2013). LFQ values were analyzed with R Bioconductor. Here, the arithmetic mean, standard deviation and relative protein abundance change in relation to the reference were calculated. Student’s *t*-test was performed for significance testing ($p < 0.05$) of single protein abundance changes.

Proteins were annotated according to the COG (clusters of orthologous groups) database (Tatusov et al., 1997) and clustered in two hierarchical levels, namely ‘metabolism’ and ‘pathway’. Groups were visualized using a color-coded circular treemap (Jahn et al., 2013). Additional, protein clusters were tested for significant changes using the R Bioconductor packages GAGE (Luo et al., 2009) and GlobalTest (Goeman et al., 2004), setting $p \leq 0.05$ and a relative fold change (FC) of 1.5 ($\log_2\text{FC} = 0.58$) as thresholds. Detailed information on gene set analysis and reasons behind the usage of GAGE and GlobalTest are given in the following section 5.5.

5.5. Transcriptome analysis

5.5.1. Sampling procedure and RNA next generation sequencing

A sample of 2 mL cultivation broth was taken directly into 4 mL of RNeasy Protect Bacteria Reagent (Qiagen GmbH, Germany), vortexed and incubated at room temperature for 5 min. Aliquots of the solution containing approximately 10^9 cells were centrifuged at $7000 \times g$ at 4°C for 10 minutes, before the supernatant was discarded and the cell pellet was shock frozen in liquid nitrogen and stored at -70°C until shipment. The samples of biological replicates were collectively shipped on dry ice for a batch RNA next generation sequencing, carried out by MFT Services (Tübingen, Germany). A detailed description about the sequencing procedure, equipment and sequence alignment is given in the appendix in section B.2 ‘Transcriptome Analysis’.

5.5.2. Statistical data analysis

After a statistical assessment of the data, the analysis of datasets from high-throughput omics-technologies, such as transcriptomics and proteomics, ultimately yields a list of differentially expressed genes or proteins. The challenge of analyzing these sometimes long lists of differential expression information lies in the extraction of meaningful and mechanistic insights to answer the scientific question that was raised.

Statistical data analysis was performed as a part of this thesis with the Bioconductor package ‘edgeR’ (Robinson et al., 2010), which was especially developed for the analysis of digital gene expression data (Robinson et al., 2007; Robinson et al., 2008).

Assessment of the list of differentially expressed genes To account for differences in sequencing depth, the raw count data was first normalized based on ‘counts per million mapped counts’ (CPM). Discrete count data as obtained by RNA-Seq was shown to follow a negative binomial distribution (McCarthy et al., 2012). Differential expression analysis was carried out following the protocol by Anders et al. (2013) using edgeR. The resulting p -values were adjusted for multiple testing according to Benjamini and Hochberg (1995) to calculate the false discovery rate (FDR). A cutoff of $FDR \leq 0.05$ was chosen to extract differentially expressed genes.

Gene set analysis Functional grouping of the individual genes into gene sets of related genes has been proven to be a useful approach to extract information about mechanistic information on the metabolic pathway level. Gene set analysis (GSA) allows a reduction of the complexity of the analysis problem from thousands of differentially expressed genes to only hundreds of pathways. The identification of differentially expressed pathways has a higher power of giving coherent results than a long list of not obviously related differentially expressed genes or proteins (Glazko et al., 2009). Furthermore, it was shown that small coordinated changes within expression of a whole pathway may have a significant biological effect, even if the changes in expression of individual genes may not be statistically significant (Subramanian et al., 2005).

A pitfall of GSA is the dependency on public availability of the pathway knowledge. Depending on the organism used, the information depth available varies substantially in repositories such as the Gene Ontology Consortium (GO) or the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al., 2000).

Three approaches to gene set analysis have been defined: the over-representation analysis (ORA), the functional class scoring (FCS) and the pathway topology based (TP) gene set analysis (Khatri et al., 2012).

ORA methods statistically correlate the significant changes of a fraction of genes within a pathway to the set of genes clustered in this pathway. The drawback of this approach is that only the counted number of genes is used and therefore every gene is treated equally, not taking fold-changes of statistical significance into account (Khatri et al., 2002). Consequently, as every gene is assumed to be independent of the other genes, the interaction of gene products in different pathways is not taken into account, and therefore the estimated significance of a pathway may be biased or in extreme cases incorrect.

FCS approaches address these three limitations by accounting also for weaker, statistically not significant individual, but coordinated changes in sets of genes assorted in pathways (Barry et al., 2005). Hereby, not only numbers of genes, but also fold-change and statistical information is included in the analysis, as well as dependencies of genes are taken into account when considering coordinated expression changes. A limitation of the method is the independent analysis of pathways, which may lead to identification of significantly changed pathways due to multiple annotations of individual genes in different pathways (Khatri et al., 2012).

If additional information on interactions of pathways is available in repositories, such as activation or inhibition, TP based methods can be applied to overcome the drawbacks of ORA and FCS. Unfortunately, this information, if available, is sparse in the case of *P. putida* KT2440. Consequently, FCS methods were applied for the analysis of transcriptome and proteome datasets.

Within the FCS methods, the significance of gene set differential expression can be calculated either based on randomization of sample labels (e.g. GlobalTest (Goeman et al., 2004)) or on a parametric gene randomization procedure, such as GAGE (Luo et al., 2009). As both tests evaluate different but related null hypothesis, a combination of the procedures achieves statistically more robust results (Tian et al., 2005; Nam et al., 2008).

5.6. Quantification of cultivations

5.6.1. Bacterial growth kinetics

A microbial cell, grown in batch cultivation, proceeds through the typical growth curve consisting of lag, acceleration, exponential, deceleration, stationary and death phases (Figure 5.3). The sequence of the growth curve is not an inherent property of the organism, but a result of its interaction with the constantly changing physico-chemical environment in which it is growing in batch cultivation (Tempest, 1970). During the lag phase, the organism adjusts its gene expression and enzyme production to its new environment. After this phase of only little or no growth, the growth rate of the organism increases until it is proliferating at its maximum rate. In this phase, growth is not limited by substrate availability and the cell population grows exponentially at a constant

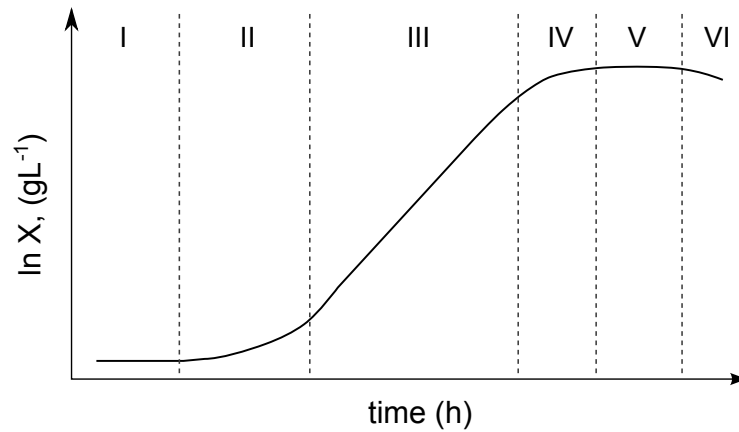


Figure 5.3.: Schematic overview of the bacterial growth curve. A microbial population typically passes through 6 stages during batch cultivation: a lag phase (I), with little or no growth, where cells adapt to their new environment and an acceleration phase (II), where the growth rate increases until the maximum growth rate is reached in the exponential phase (III). The population grows at constant maximal growth rate until the substrate becomes limiting and growth rate decelerates (IV) until the substrate is completely depleted and the population enters the stationary phase (V), finally entering the death phase (VI).

maximum growth rate μ_{\max} (in h^{-1}). As soon as substrate concentration becomes limiting, growth speed decelerates until the substrate is completely consumed in the stationary phase.

5.6.2. Mass balances

Mass balances offer valuable information on reaction rates, such as biomass formation, substrate uptake and product formation rates.

Reaction parameters in batch cultivation

Calculation of the specific growth rate μ In a closed batch cultivation system, the growth rate μ can be derived from the biomass balance:

$$\frac{dm_x}{dt} = \mu \cdot c_x \cdot V_R \quad (5.1)$$

Here, m_x is the biomass in g, c_x the biomass density in gL^{-1} and V_R the cultivation volume in L. In a batch cultivation, the reaction volume V_R is assumed to be constant. Therefore, the growth rate equals

$$\mu = \frac{1}{c_x} \frac{dc_x}{dt} \quad (5.2)$$

The doubling time t_d (in h), which also resembles the generation time τ , can be directly derived by integration of Eq. 5.2

$$\tau = \frac{\ln 2}{\mu} \quad (5.3)$$

The growth rate of an organism is dependent on the nutrient availability, assuming one limiting substrate, as formulated by Monod (1949):

$$\mu = \mu_{max} \cdot \frac{c_s}{c_s + K_s} \quad (5.4)$$

Here, c_s is the substrate concentration of the limiting substrate in gL^{-1} , μ_{max} the maximum specific growth rate in h^{-1} and K_s is the limiting substrate concentration at which the specific growth rate is half its maximum value. Notably, the K_s value in the Monod model does not exactly represent the saturation constant for substrate uptake, but only an overall saturation constant for the whole growth process. However, K_s values mostly do not differ significantly from K_m values of the enzymes involved in substrate uptake, because substrate uptake is often closely connected to the control of substrate metabolism (Villadsen et al., 2011).

During the exponential growth phase of a batch cultivation it can be assumed that $K_s \ll c_s$. Even though no specific data is available for *P. putida* KT2440, K_s values for glucose were found to be in the range of 4 – 150 mgL^{-1} for *E. coli* and *Saccharomyces cerevisiae*, respectively (Villadsen et al., 2011). Therefore, during exponential growth the following equation is valid

$$\mu = \mu_{max} \quad (5.5)$$

The maximum specific growth rate was calculated as linear regression and least squares fitting of $\ln(c_x)$ over time during the exponential growth phase of the population ($\mu = \mu_{max} = \text{const.}$).

Calculation of specific substrate uptake q_s and production rates q_p Massbalances for substrate (s) and product (p) can be set up in the same way as described for biomass (x) (Eq. 5.1):

$$\frac{dm_s}{dt} = -q_s \cdot c_x \cdot V_R \quad (5.6)$$

$$\frac{dm_p}{dt} = q_p \cdot c_x \cdot V_R \quad (5.7)$$

Here, m_s and m_p are the masses of the substrate and the product in g. q_s and q_p are the biomass specific substrate and production rates in $\text{gg}^{-1}\text{L}^{-1}$.

Reaction parameters in continuous cultivation

As introduced in section 3.2, at steady-state mode, the flow rate into the reactor equals the flow rate out of the reactor ($F_{in} = F_{out}$). Therefore, the reaction volume is assumed to stay constant ($V_R/dt = 0$). The ratio of the flow rate to the reaction volume is defined as dilution rate D :

$$D = \frac{F}{V_R} \quad (5.8)$$

Furthermore, at steady state, no net mass accumulation occurs. Consequently, the mass of the compound produced by the reaction is equal to the difference in mass of the compound between the liquid feed and the outlet of the reactor. Considering biomass, no biomass is present in the liquid feed. The biomass balance can be derived as follows:

$$\frac{dm_x}{dt} = V_R \cdot \frac{dc_x}{dt} = \mu \cdot c_x \cdot V_R - F \cdot c_x \quad (5.9)$$

$$\frac{dc_x}{dt} = \mu \cdot c_x - D \cdot c_x \quad (5.10)$$

Equivalent to the biomass balance at steady state in a chemostat, mass balances for substrate and products can be defined as:

$$\frac{dc_s}{dt} = D \cdot (c_{s0} - c_s) - q_s \cdot c_x \quad (5.11)$$

$$\frac{dc_p}{dt} = D \cdot (c_{p0} - c_p) + q_p \cdot c_x \quad (5.12)$$

Here, c_{i0} is the concentration of compound i in the liquid feed in g L^{-1} .

Yield coefficients Scaling any particular rate q_i with another rate q_j is resulting in the yield coefficient $Y_{i/j}$:

$$Y_{i/j} = \frac{q_i}{q_j} \quad (5.13)$$

In this thesis, different yield coefficients were calculated, ranging from yield of biomass per substrate ($Y_{X/S}$) to yield of product per biomass ($Y_{P/X}$).

Respiration rates and exhaust gas analysis With a gaseous substrate or product, the mass balances need to be modified by considering a rate of transfer from the gas phase to the liquid medium.

The oxygen transfer rate (OTR, in $\text{mol L}^{-1} \text{h}^{-1}$) is proportional to the mass transfer coefficient $k_L a$

and the concentration driving force for mass transfer:

$$OTR = k_L a (c_{O_2}^* - c_{O_2}) \quad (5.14)$$

Here, $c_{O_2}^*$ (in molL^{-1}) is the oxygen concentration in the liquid which is in equilibrium with the gas phase (gas-liquid interphase), whereas c_{O_2} (in molL^{-1}) is the oxygen concentration in the bulk liquid.

At steady state conditions, since no oxygen is accumulating, the OTR must equal the oxygen uptake rate (OUR, in $\text{molL}^{-1}\text{h}^{-1}$):

$$\frac{dc_{O_2}}{dt} = OTR - OUR = 0 \quad (5.15)$$

The OUR is defined as the difference of the amount of oxygen of the inlet ($n_{O_2,in}$, molh^{-1}) and exhaust gas flow ($n_{O_2,out}$, molh^{-1}) per working volume V_R :

$$OUR = \frac{n_{O_2,in} - n_{O_2,out}}{V_R} \quad (5.16)$$

Connecting Eq. 5.16 with the ideal gas law (Eq. 5.17) and assuming a valid nitrogen inert gas balance (Eq. 5.18) and isobaric and isothermal inlet and outlet air flow conditions, the OUR can be calculated according to Eq. 5.19:

$$p\dot{V}_g = \dot{n}RT \quad (5.17)$$

$$\dot{n}_{N_2,in} = \dot{n}_{N_2,out} \quad (5.18)$$

$$OUR = \frac{p}{RT} \frac{\dot{V}_{g,in}}{V_R} \left(y_{O_2,in} - y_{O_2,out} \left[\frac{1 - y_{O_2,in} - y_{CO_2,in}}{1 - y_{O_2,out} - y_{CO_2,out}} \right] \right) \quad (5.19)$$

Here, \dot{V}_g (in Lh^{-1}) is the volumetric gas flow rate, R (in $\text{Jmol}^{-1}\text{K}^{-1}$) the universal gas constant, T (in K) the absolute temperature, p (in Pa) the pressure and y_i (in %) the volumetric gas fraction of gas i .

The carbon dioxide evolution rate CER (in $\text{molL}^{-1}\text{h}^{-1}$) can be formulated accordingly:

$$CER = \frac{p}{RT} \frac{\dot{V}_{g,in}}{V_R} \left(\left[\frac{1 - y_{O_2,in} - y_{CO_2,in}}{1 - y_{O_2,out} - y_{CO_2,out}} \right] y_{CO_2,out} - y_{CO_2,in} \right) \quad (5.20)$$

The dimensionless respiratory quotient RQ can be derived by

$$RQ = \frac{CER}{OUR} \quad (5.21)$$

5.6.3. Carbon balancing

Carbon balances for all processes were set up considering biomass formation, CO₂ evolution and the concentration of residual glucose in the culture medium to check for obvious errors in analytical measurements and for possible side-product formation. The average elementary cell composition of bacteria CH_{1.8}O_{0.5}N_{0.2} (Villadsen et al., 2011) was used, as no elementary cell composition of *P. putida* KT2440 under carbon limited conditions is available.

The carbon balances were calculated as follows:

$$a \cdot CH_2O - b \cdot CH_{1.8}O_{0.5} - c \cdot CO_2 - d \cdot CH_2O_{residual} \stackrel{!}{=} 0 \quad (5.22)$$

$$(5.23)$$

The coefficients $a - d$ represent the measured concentration of the specific compound (in C·molL⁻¹).

5.6.4. Maintenance demands

Maintenance demands on glucose (m_s , in g_{GLC}g_{CDW}⁻¹h⁻¹) were calculated by following the Pirt's equation (Pirt, 1965):

$$q_s = m_s + \mu/Y_{X/S_{true}} \quad (5.24)$$

where q_s is the specific rate of glucose consumption (in g_{GLC}g_{CDW}⁻¹h⁻¹), μ is the specific growth rate (in h⁻¹), and $Y_{X/S_{true}}$ is the true yield of biomass on glucose (in g_{CDW}g_{GLC}⁻¹). Detailed information about the calculation procedure can be found in the appendix in section C.2 'Calculation of maintenance demands'.

5.6.5. Propagation of uncertainties

Gaussian error propagation was used to calculate the standard deviation σ_f of parameters, that were a function of at least two individually measured variables (f(x,y)). Assuming uncorrelated uncertainties, σ_f was determined from the uncertainties of the variables (σ_x and σ_y) which propagate to their combination in the function:

$$\sigma_f = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial f}{\partial y}\right)^2 \sigma_y^2} \quad (5.25)$$

CHAPTER 6

THE CELL CYCLE AS ORIGIN OF POPULATION DYNAMICS

This chapter contains the results and the discussion of the investigation of the cell cycle as a biological factor causing population heterogeneity. Parts of this chapter have been published as ‘Subpopulation-proteomics reveal growth rate, but not cell cycling, as a major impact on protein composition in *Pseudomonas putida* KT2440’¹

Physiological differences of individual cells within a clonal cell population are a commonly accepted fact (Avery, 2006; Müller et al., 2010). Nevertheless, their appearance and impact on process performance still remains rather unclear. Among many proposed factors (see section 3.1, Figure 3.2), cell cycling is one of the suggested drivers of heterogeneity (Avery, 2006; Müller et al., 2010). Fueled by the finding of Ackermann et al. (1995) that PHA accumulated in dependency on the chromosome content in *Methylobacterium rhodesianum*, it was discussed that the biosynthesis of compounds of biotechnological interest might be dependent on the cell cycle phase (Müller et al., 2010). Population heterogeneity caused by cell cycling could therefore have significant impact on the overall process performance (Lencastre Fernandes et al., 2011).

In this chapter, we investigated if the protein inventory of a cell is dependent on the cell cycle phase. We focused on the question if subpopulations growing at the same growth rate, but being in different phases of the cell cycle, were different from each other at the level of their protein content. Furthermore, we wanted to know if the subpopulation composition differs dependent on specific growth rates, e.g. whether slow growing cells with longer cell cycling phases might specialize between proliferation and production phases, while subpopulations arising at faster growth rates might invest into different protein species.

¹Sarah Lieder, Michael Jahn, Jana Seifert, Martin von Bergen, Susann Müller, Ralf Takors (2014) Applied Microbiology and Biotechnology Express 4:71 (Appendix A)

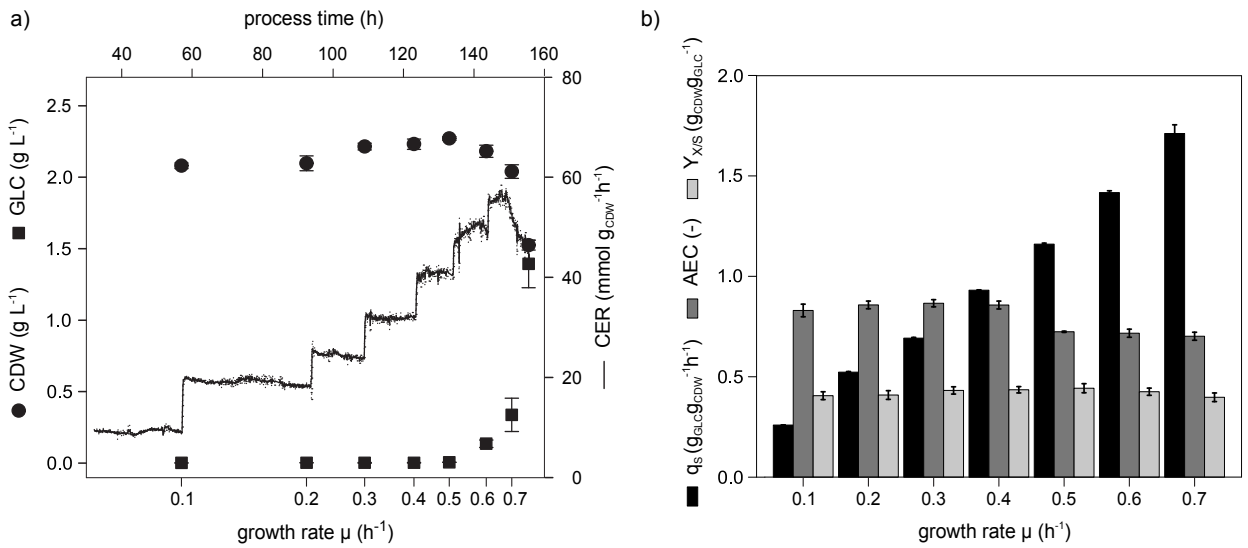


Figure 6.1.: Physiological data of *P. putida* KT2440 continuous cultivations at different growth rates. The growth rate was stepwise increased until a wash-out of the population was monitored (a). The cell dry weight (CDW, g L⁻¹, black dots) and the residual glucose concentration (GLC, g L⁻¹, black squares) were measured after 5 residence times of one specific growth rate μ (h⁻¹) at steady-state. The carbon dioxide emission rate CER (mmol L⁻¹h⁻¹, black line) was monitored online. Error bars and lines (CER, grey dotted line) represent the standard deviation of independent biological triplicates. Specific glucose uptake rates q_s (g_{GLC}g_{CDW}⁻¹h⁻¹, black bars), the adenylate energy charge (AEC, grey bars) and the biomass yield coefficient $Y_{X/S}$ (g_{CDW}g_{GLC}⁻¹, light grey bars) were calculated for each specific growth rate (b).

6.1. Design of the experimental set-up

A careful choice of the experimental set-up is crucial for dissecting the impact of the cell cycle from the variety of other parameters influencing population dynamics, which overlay, interact or even amplify each other. Here, we applied continuous cultivations (chemostats). Contrary to batch cultivations, in which cells are growing at differing growth rates due to constantly changing cultivation conditions (Unthan et al., 2014), chemostat cultivations provide a controlled environment at a constant growth rate. Consequently, they allow the investigation of the influence of one specific parameter, while all other cultivation parameters are kept constant.

As a first step, a chemostat was set up. The growth rate –set externally by selecting a defined dilution rate– was step-wise increased until cells could not reproduce fast enough to keep the population density constant and therefore, were washed out (Figure 6.1a). All samples were taken under steady-state conditions after at least 5 residence times of one constant growth rate and a stable carbon dioxide emission rate (CER).

6.2. Physiological characterization of the average population

Physiological parameters and the energetic state of the averaged cell population were determined in order to build a basis for the interpretation and comparison of subpopulation and proteome investigations. Here, the biomass yield ($Y_{X/S}$), the biomass specific glucose uptake rate (q_s) and the adenylate energy charge (AEC) were calculated at slow to fast growth rates ($0.1 < \mu < 0.7 \text{ h}^{-1}$, Figure 6.1b). Between $0.1 < \mu < 0.5 \text{ h}^{-1}$ an increase of growth rate resulted in a gradual increase of $Y_{X/S}$ by 10%. Further acceleration of growth resulted in yield reductions, returning to the initial $Y_{X/S}$ of $\mu = 0.1 \text{ h}^{-1}$ at $\mu = 0.7 \text{ h}^{-1}$ (−10%). The energetic state of the cell population was analyzed via the AEC, which indicates the relative saturation of high-energy phospho-anhydride bonds available in the adenylate pool of the cell. The AEC remained constant with increasing growth rate until $\mu = 0.5 \text{ h}^{-1}$. Further increase of the growth rate resulted in a reduction of the AEC level by −18% (p -value < 0.01). q_s was increasing linearly with increasing growth rate.

6.3. Quantification of subpopulations via flow cytometry

(Sub-)Population dynamics were analyzed via flow cytometry. A representative dataset of the distribution of forward scatter (FSC) and DAPI fluorescence, plotted as histograms, can be found in Figure 6.2 (2).

In a first analysis step, non-cell particles were excluded from the cell population via FSC gating. Considering the FSC dataset, no clear subpopulations could be identified. Nonetheless, it could be observed that the average FSC signal increased with increasing growth rates.

In a second step, the distribution of DAPI fluorescence within the cell population was analyzed. DAPI is a fluorescence marker that specifically labels A/T-rich regions of DNA. It was found to be a highly selective and stable marker for quantification of the DNA content of cells (Müller et al., 2010). Contrary to the FSC signals, clear subpopulations exhibiting a certain amount of DAPI fluorescence could be identified. For example, the first peak of the DAPI fluorescence histogram can be interpreted as the number of cells containing one chromosome equivalent (C1). Consequently, the second peak consists of cells carrying a double chromosome equivalent content (C2). For *P. putida* KT2440 grown under minimal media conditions, it was shown that the first DAPI peak actually refers to a single chromosome content (Jahn et al., 2014). Therefore, it is possible to interpret the chromosome content quantitatively, instead of only referring to ‘chromosome equivalents’. At faster growth rates (Figure 6.2 (2), $\mu = 0.7 \text{ h}^{-1}$), a third subpopulation of cells containing more than two chromosomes (Cx) could be detected.

With the chromosome content in hand, it is possible to relate the subpopulations to the bacterial cell cycle and to assign a cell cycle phase to each subpopulation (refer to Figure 3.3). Cells

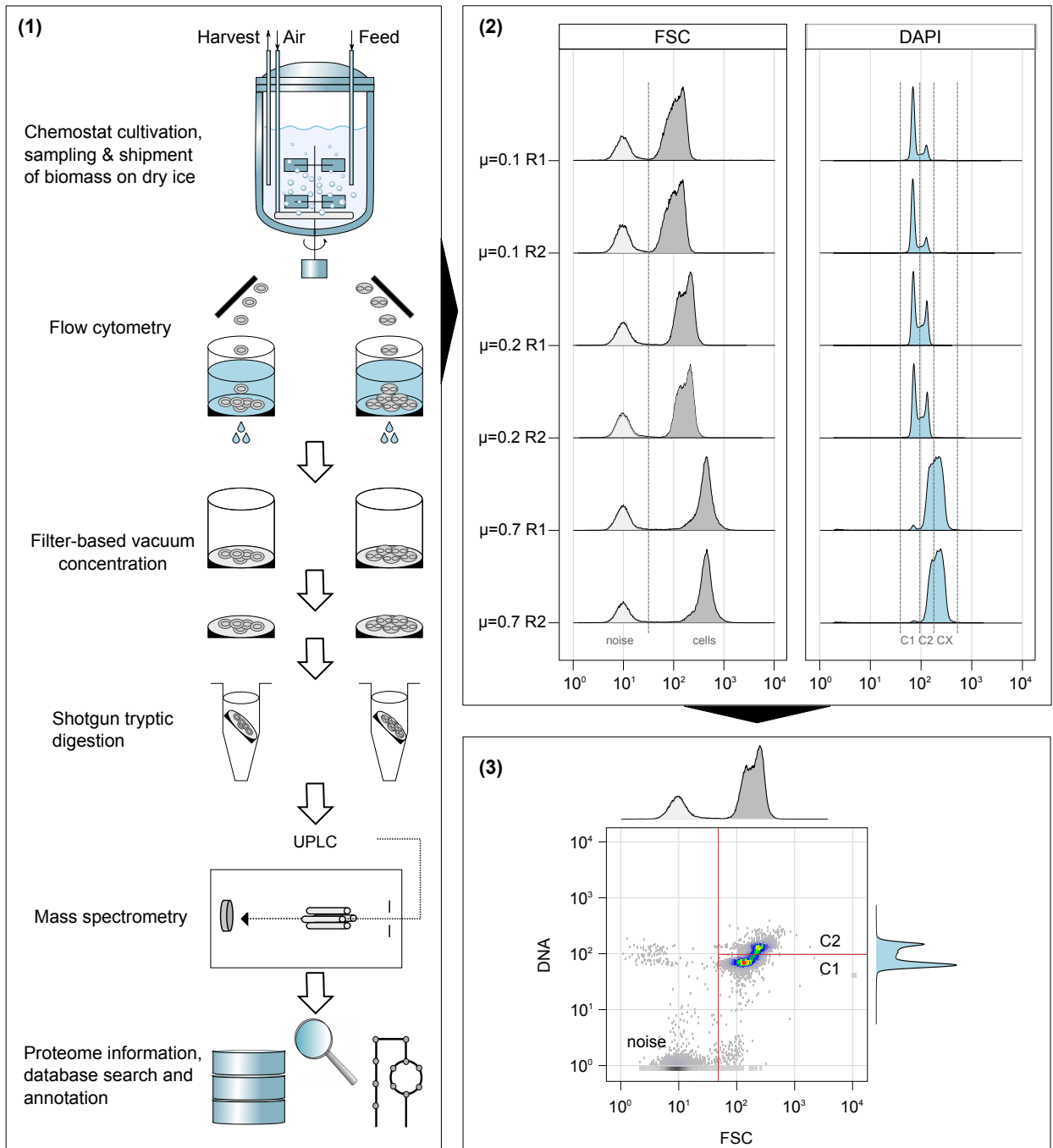


Figure 6.2.: DAPI fluorescence (DAPI) and forward scatter (FSC) as parameters for cell sorting. (1) Schematic overview of the workflow applied from continuous cultivation over cell sorting on filter wells and tryptic digestion to label-free mass spectrometry for subpopulation proteomics. (2) Fluorescence of DAPI (light blue) and FSC (grey) of *P. putida* KT2440 were measured by flow cytometry. (3) To detect changes in protein abundance dependent on the cell cycle stage and the growth rate, cells harvested at steady state conditions at 3 different growth rates ($\mu = 0.1 \text{ h}^{-1}$, $\mu = 0.2 \text{ h}^{-1}$, $\mu = 0.7 \text{ h}^{-1}$, in two biological replicate cultivations (R1, R2)), were sorted based on fluorescence and forward scatter. Gates (red lines) were chosen to exclude technical noise and to sort cells into three subpopulations C1, C2 and CX, depending on the strength of the DAPI fluorescence (adapted from Jahn et al. (2013)).

containing a single chromosome have just divided, but did not start replicating yet (B phase), whereas cells with a double chromosome content just finished replication, but did not divide yet (pre-D/D phase). The Cx subpopulation can be interpreted as cells growing with an uncoupled cell cycle, maintaining a fast growth rate.

The DNA content was identified as the major differential parameter between subpopulations. DAPI staining does not only allow a ‘yes’ or ‘no’ marker decision, but rather a quantitative subpopulation determination, which can be precisely related to the cell cycle. The DAPI fluorescence signal was chosen as a selection marker for subsequent proteome analysis.

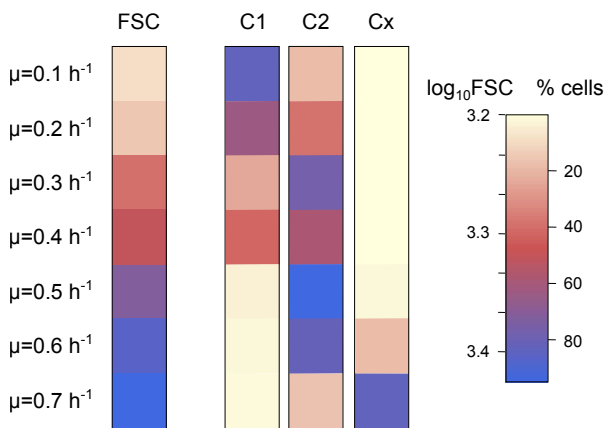


Figure 6.3.: Subpopulation distributions at different growth rates. The average forward scattering (FSC, in arbitrary fluorescence units, $\log_{10}\text{FSC}$) and the percentage of cells containing one (C1), two (C2) or more than two (Cx) chromosomes, as determined by flow cytometry, are depicted as color-coded heatmap.

The population composition with respect to DNA content was altered as a function of growth rates (Figure 6.3). At $\mu = 0.1 \text{ h}^{-1}$, the majority of cells, $82.0 \pm 0.3\%$, contained a single chromosome, while only $18.0 \pm 0.2\%$ contained a double chromosome. No cells containing more than two chromosomes could be detected. With increasing growth rate the fraction of C1 decreased, while the fraction of C2 increased, until at $\mu = 0.7 \text{ h}^{-1}$ only $1.4 \pm 0.8\%$ of the population belonged to the C1 subpopulation, while $16.1 \pm 0.1\%$ of cells contained a double chromosome content and $82.5 \pm 1.0\%$ showed a more than double chromosome content.

6.4. Subpopulation proteome analysis

The analysis of the protein content of cell cycle subpopulations required a controlled and reliable workflow (Jahn et al., 2013), which is summarized in Figure 6.2 (1). Cells were sorted at three growth rates (0.1 h^{-1} , 0.2 h^{-1} and 0.7 h^{-1}) according to their chromosome content (C1, C2 and Cx) and differences between the subpopulation proteome profiles were assessed as a basis of their phenotypes. Fold changes of protein abundance were calculated in relation to the reference population ($\mu = 0.2 \text{ h}^{-1}$). The reference population was sorted in order to exclude influences of the sorting procedure on the protein content and unsorted cells of the 0.2 h^{-1} grown population were used as an unaffected control population.

In total, 677 unique proteins (annotated and hypothetical) could be detected, whereof 351 proteins were found in at least one replicate of all subpopulations and 245 proteins were found across all replicates. 707 different functions of 677 unique proteins were annotated using the database of

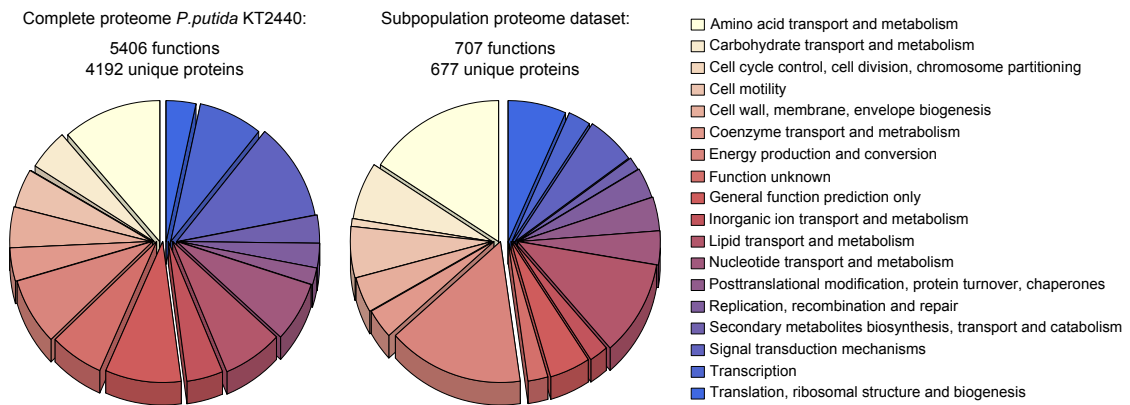


Figure 6.4.: Visualization of COG annotation. Proteins were grouped into 18 functional classes using COG (Tatusov et al., 1997). The total number of assigned proteins functions and the number of unique proteins that are annotated for the complete proteome of *P. putida* KT2440 (left) are contrasted with the subpopulation proteome dataset (right).

clusters of orthologous groups (COG) (Tatusov et al., 1997). A comparison between the COG annotation of *P. putida* KT2440 and the subpopulation proteome dataset can be found in Figure 6.4. All functional groups were represented in the subpopulation dataset. Furthermore, 98.2% of the proteome of the control population could be found in the reference population proteome without significant changes, indicating only a small influence of cell sorting on protein recovery and confirming the quality of the analysis (data not shown).

Changes in protein abundance were declared to be significant if they exceeded a 1.5 fold change (FC) and showed statistical significance (p -value < 0.05). Gene set analysis methods were used to detect changes in metabolic pathways, applying the same significance filter as for individual proteins (Luo et al., 2009; Goeman et al., 2004).

Comparing cell cycle subpopulations at the same growth rate, no changes in metabolic pathways could be observed. Looking at the level of individual proteins, only little significant changes were observable (Figure 6.5a). At $\mu = 0.1 \text{ h}^{-1}$ and $\mu = 0.7 \text{ h}^{-1}$, only three out of all proteins detected had significantly altered levels, among them the cell division protein FtsZ. Its abundance was found to be 3.6 fold lower in the C1 subpopulation as compared to the C2 subpopulation at $\mu = 0.1 \text{ h}^{-1}$. FtsZ is a bacterial tubulin homologue, self-assembling into a ring at mid-cell level and localizing the bacterial divisome machinery (Adams et al., 2009; Weart et al., 2007). The other significantly changed proteins could not be directly linked to the cell cycle or connected to any other unique functional group or metabolic pathway (details can be found in section A.3 in the appendix).

These marginal changes of the proteome in dependency of the cell cycle were surprising, since cell cycle dependent periodic gene expression has been reported for many organisms (Wittenberg et al., 2005; Rustici et al., 2004; Laub et al., 2000). Considering the protein coverage of more than one third of the annotated proteins in the functional group ‘cell cycle’ in our study, the lack of abundance changes cannot only be attributed to protein coverage. Another assumption that

could explain our results would be an activity regulation on a different than a translational level, e.g. on a posttranslational level. Recently, Waldbauer et al. (2012) reported similar findings of only small proteome changes during the cell cycle of the cyanobacterium *Prochlorococcus* compared to extensive changes in the transcriptome, strengthening the validity of our observations and implications.

Comparing cell cycle subpopulations at different growth rates, major changes in metabolic pathways could be detected (Figure 6.5b and c). Slow growing cells of both subpopulations, C1 and C2, showed higher abundance of proteins annotated in the functional groups ‘cell motility’, while proteins involved in ‘cell cycle control, cell division and chromosome partitioning’ (cell cycle) were additionally highly abundant in subpopulation C2. Regarding ‘cell motility’, four main chemotaxis signaling proteins (CheA ($\log_2\text{FC}=2.2$, PP_4338), CheB ($\log_2\text{FC}=3.7$, PP_4337), CheW ($\log_2\text{FC}=3$, PP_4332) and CheV ($\log_2\text{FC}=3.2$, PP_2128)) and 6 methyl accepting chemotaxis transducers were found in higher abundance, anticipating increased motility and chemotaxis response at slow growth rates. Moreover, a significant increase of poly(3-hydroxyalkanoate) synthetases PhaA ($\log_2\text{FC}=3$, PP_5003) and PhaC ($\log_2\text{FC}=4.5$, PP_5005) could be detected, indicating higher PHA production at slow compared to fast growth rates in the chemostat.

Chemotaxis and cellular motility are well known responses to nutrient-poor conditions in natural environments (Harshey, 2003; Soutourina et al., 2003). The observations of our proteome analysis of the slowly growing subpopulations are in agreement with findings of transcriptome studies in ‘average populations’ of other species. Nahku et al. (2010) showed in *E. coli*, that genes involved in motility were over expressed at slower growth rates in direct comparison to faster growth conditions. Moreover, in accordance to our finding, chemostat studies in *P. oleovorans* reported a higher PHA productivity at slow in comparison to faster growth rates (Preusting et al., 1993).

When looking at the fast growing population, both subpopulations, C2 and Cx, showed a high abundance of proteins annotated in the pathway ‘Translation, ribosomal structure and biogenesis’ (Translation), while proteins of ‘Signal transduction mechanisms’ (Signaling) and ‘Lipid transport and metabolism’ (Lipids) were significantly less abundant. Reflecting the accelerated protein synthesis associated with faster growth, 11 tRNA synthetases and 25 ribosomal proteins showed significantly higher abundance (for details, refer to Figure A.5 in section A.3 in the appendix). The observation, that cells at higher growth rates increasingly invest into translation machinery and protein biosynthesis, is also in agreement with observations in eukaryotes like *S. cerevisiae* (Reb-negger et al., 2014) and prokaryotes such as *Salmonella typhimurium* (Schaechter et al., 1958). Moreover, proteins of typical carbon storage pathways e.g. PHA synthesis were found in lower abundance in the fast growing subpopulations. Noteworthy, the seemingly lower abundance of proteins connected to the ‘Cell Cycle’ (C2 versus Cx) was mainly due to the single protein change of the poorly characterized PP_3128 and was therefore neglected.

Surprisingly, an increase of growth rate was not mirrored by major changes among proteins involved

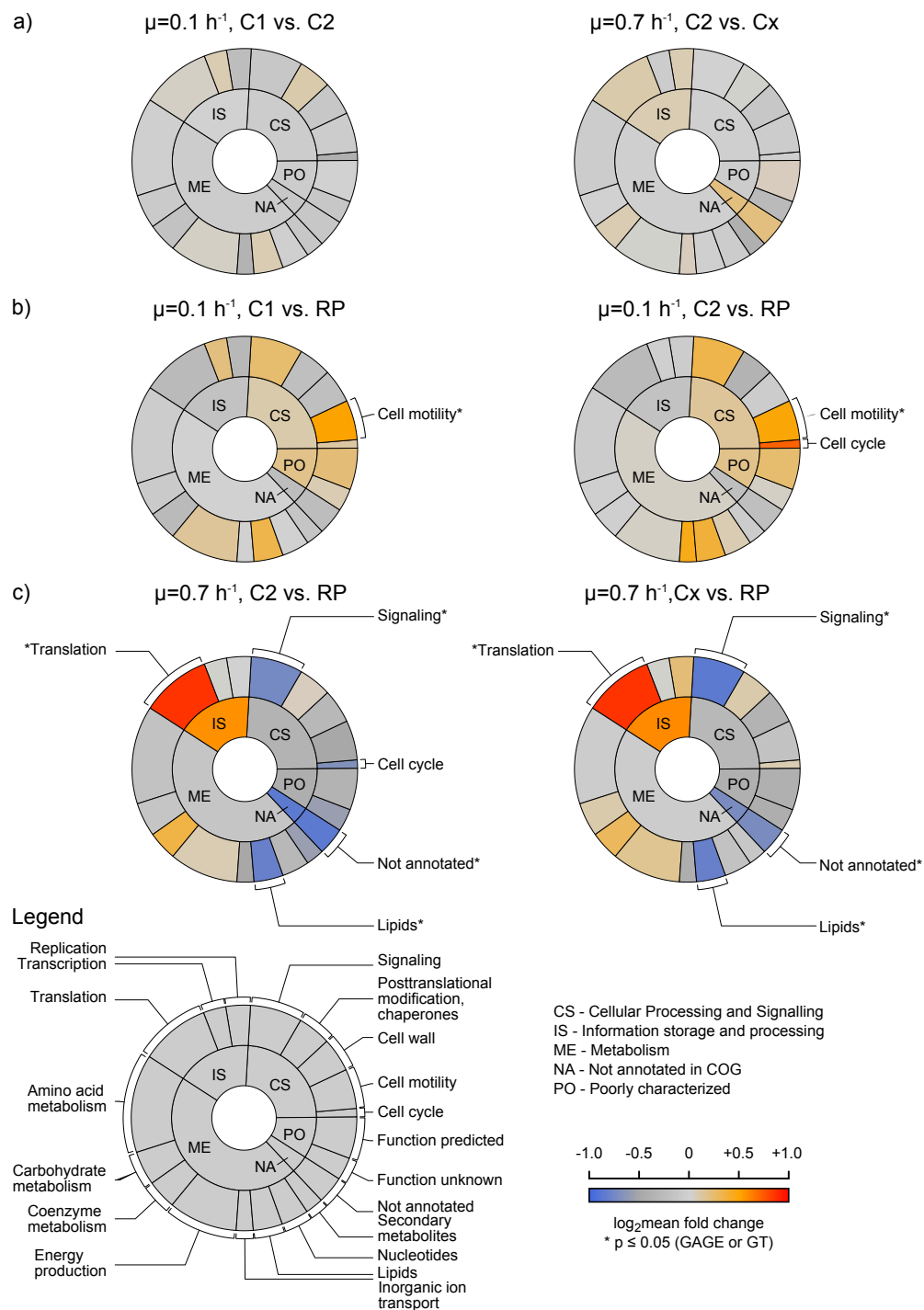


Figure 6.5.: Circular treemaps visualizing differentially expressed functional protein categories. Proteins detected by mass spectrometry were clustered according to their pathway annotation in COG covering two levels of specificity (Tatusov et al., 1997). The size of a sector is proportional to the number of proteins found in one specific pathway in relation to the total protein number. The color code represents the \log_2 mean fold change (\log_2 FC) of protein quantity in one pathway. The blue color indicates an under representation and the color red an over representation of the proteins in a pathway compared to the reference population (RP, $\mu = 0.2 \text{ h}^{-1}$). Pathways with a fold change in the range $\log_2 \text{FC} < -0.58$ and $\log_2 \text{FC} > 0.58$ are labeled with the respective pathway name. Pathways that were significantly changed using GAGE (Luo et al., 2009) and Globaltest (Goeman et al., 2004) gene set analysis are additionally marked (*). A. Comparison of the subpopulations C1/C2 and C2/Cx at growth rates 0.1 h^{-1} and 0.7 h^{-1} . B. Comparison of the subpopulations C1 and C2 at $\mu = 0.1 \text{ h}^{-1}$ with RP. C. Comparison of the subpopulations C2 and Cx at $\mu = 0.7 \text{ h}^{-1}$ with RP. This figure has been published in Lieder et al. (2014).

in carbohydrate and energy metabolism, irrespective of the almost 6.5-fold increase of the specific glucose uptake rate (Figure 6.1). When interpreting the subpopulation proteome dataset, one has to keep in mind that relative changes of protein abundance and not absolute quantity changes were measured. Growth rate dependent absolute changes of protein quantity for ‘average’ cells were first elucidated by Schaechter et al. (1958). This pioneering study revealed an exponential dependency of protein, DNA and RNA contents and therefore, cell size, while increasing the growth rate (Maaløe et al., 1966; Schaechter et al., 1958; Bremer et al., 2004). Here, we acquired relative information on the cell size via the FSC signal. In accordance to various other studies, the FSC increased with increasing growth rates (Skarstad et al., 1985; Hewitt et al., 1999; Neumeier et al., 2013) (Figure 6.3). Following the rationale of Schaechter et al. (1958), increasing cell size can be interpreted as an increase of protein content per cell. We assume that the glucose uptake is proportional to the elevated production of proteins at high growth rates, thus increasing the absolute protein quantity but leaving the relative quantity unchanged.

The abundance of proteins inside a cell cannot readily be translated into protein activity. The additional assessment of ‘sub-metabolomes’ could give deeper insights into the metabolic activity of the cell in dependence on the cell cycle. However, measuring the metabolome in subpopulations of prokaryotes is still coming of age (Zenobi, 2013). High turnover rates of metabolites are directly connected to difficulties of a reliable fixation of the pools during sampling and the cell sorting procedure. Here, sampling methods have to be especially developed that allow quenching, minimization of metabolite leakage and to keep the cell walls intact for cell sorting. Improvements of the analytical method are needed to lower detection limits, increase coverage and allow better and faster identification of metabolites while reducing the sorting time and problems related to fixation. Therefore, until today, ‘sub-proteome’ snapshots are the method of choice, with proteins being stable and allowing insights into complex protein expression patterns that reveal deep functional information.

In summary, the results of the subpopulation proteome analysis revealed almost identical protein composition of cells differing in DNA content but with identical growth rate, whereas the proteome of cells cultivated at different growth rates showed significant differences in specific pathways.

Investigating the impact of the cell cycle on population heterogeneity in the operation mode ‘chemostat’ allowed a view on subpopulation physiology without superimposition of impacts of the growth rate and the cell cycle. To our surprise, the proteome data did not hint towards the physiological specialization of cells in different cell cycle stages. The discussed hypothesis of shared tasks of subpopulations in B and pre-D/D phase for e.g. carbon storage or protein production/growth could not be supported by proteome analysis. This result is surprising, since subpopulations sorted according to their DNA content appear to be physiologically highly similar at same growth rates. Surely, the proteome of subpopulations cannot be equated with single cell proteome compositions. Nevertheless, the high resemblance of the subpopulation protein pattern, regardless of the growth

rate investigated, points towards their nearly identical physiological behavior. The results give rise to the assumption that the cell cycle itself has a minor impact on population heterogeneity under the conditions tested.

CHAPTER 7

THE ENVIRONMENTAL CONDITION AS ORIGIN OF POPULATION DYNAMICS

This chapter contains the results and the discussion of the investigation of the environmental condition as an external factor causing population heterogeneity. Parts of this chapter have been published as ‘Environmental stress speeds up DNA replication in *Pseudomonas putida* in chemostat cultivations.’¹

Industrial fermentations provide challenging environmental conditions for the microbial cell population (Schweder et al., 1999). Limited mass transfer and mixing in large scale cultivations cause significant local gradients of e.g. oxygen availability or substrates inside the bioreactor. A cell, which is circulating through the bioreactor, is exposed to continuously changing environmental conditions and consequently has to adjust its physiology constantly (Fritzsche et al., 2012; Enfors et al., 2001). It was shown that challenging industrial growth conditions can lead to undesired phenotypes, including subpopulations with reduced or even stopped product formation capacities (Lara et al., 2006; Enfors et al., 2001; Lencastre Fernandes et al., 2011; Carlquist et al., 2012).

Until now, the underlying mechanisms of population split-up caused by industrially relevant stress conditions remain mostly obscure. In this chapter we aim at investigating population dynamics that result from stressful environmental conditions, typically occurring in large-scale fermentation set-ups.

¹Sarah Lieder, Michael Jahn, Joachim Koepff, Susann Müller, Ralf Takors (2016) *Biotechnology Journal* 11(1):155-63 (Appendix B)

7.1. Design of the experimental set-up

A thought-through experimental set-up is essential for the investigation of population dynamics, as reasoned in section 6.1. Three different examples of industrially-relevant stressful environments were chosen: (i) *decreased iron availability*, a typical example of error-prone large-scale media composition and bio-availability, (ii) *oxygen deprivation*, a common problem in large-scale cultivations due to limited mass transfer and of utmost importance regarding strictly aerobic *P. putida* strains and (iii) *solvent exposure*, a stress factor occurring in two-phase biocatalytic cultivation systems, which are typical industrial set-ups for *P. putida* based processes.

Continuous steady-state cultivations were used to specifically compare equally fast growing cells under non-stressed and stressed conditions. As mentioned before, chemostat studies offer the inherent advantage to prevent superimposing signals on population distributions usually occurring in batch experiments (for further information refer to section 3.2) (Skarstad et al., 1985; Wiacek et al., 2006; Müller et al., 2003; Carlquist et al., 2012).

Under non-stressed cultivation conditions, all nutrients were supplied in excess, except glucose (carbon limitation). The growth rate was stepwise increased until the maximum growth rate of *P. putida* KT2440 was reached, resulting in the wash-out of the population (??a).

Under stressed cultivation conditions, cells were grown at a constant growth rate of $\mu = 0.2 \text{ h}^{-1}$. The cultivation was started under reference conditions (non-stressed). A stress-shift was introduced after 5 residence times and the stress environment was kept constant until cells had adapted to the new conditions (showing steady-state growth), notably at the same growth rate of $\mu = 0.2 \text{ h}^{-1}$. In the end of the cultivation, the cells were shifted back to reference conditions, in order to observe whether the population showed identical physiological features as before (??b).

7.2. Quantification of subpopulations via flow cytometry

Samples for flow cytometry analysis were taken at steady-state conditions and the forward scatter (FSC), the side scatter (SSC) and the DNA content of the cells (DAPI) were analyzed. As already observed in chapter 6, the DNA content analysis showed clearly distinguishable subpopulations, also under stressed conditions (see ??). Three subgroups of cells were identified and allocated to the specific cell cycle phases, as already described in detail in section 6.3:

- (i) subpopulation C1 with a single chromosome, representing cells in B phase that just divided and did not start to replicate their DNA yet
- (ii) subpopulation C2 containing two chromosomes, representing cells in pre-D or D phase that finished replication but did not divide yet

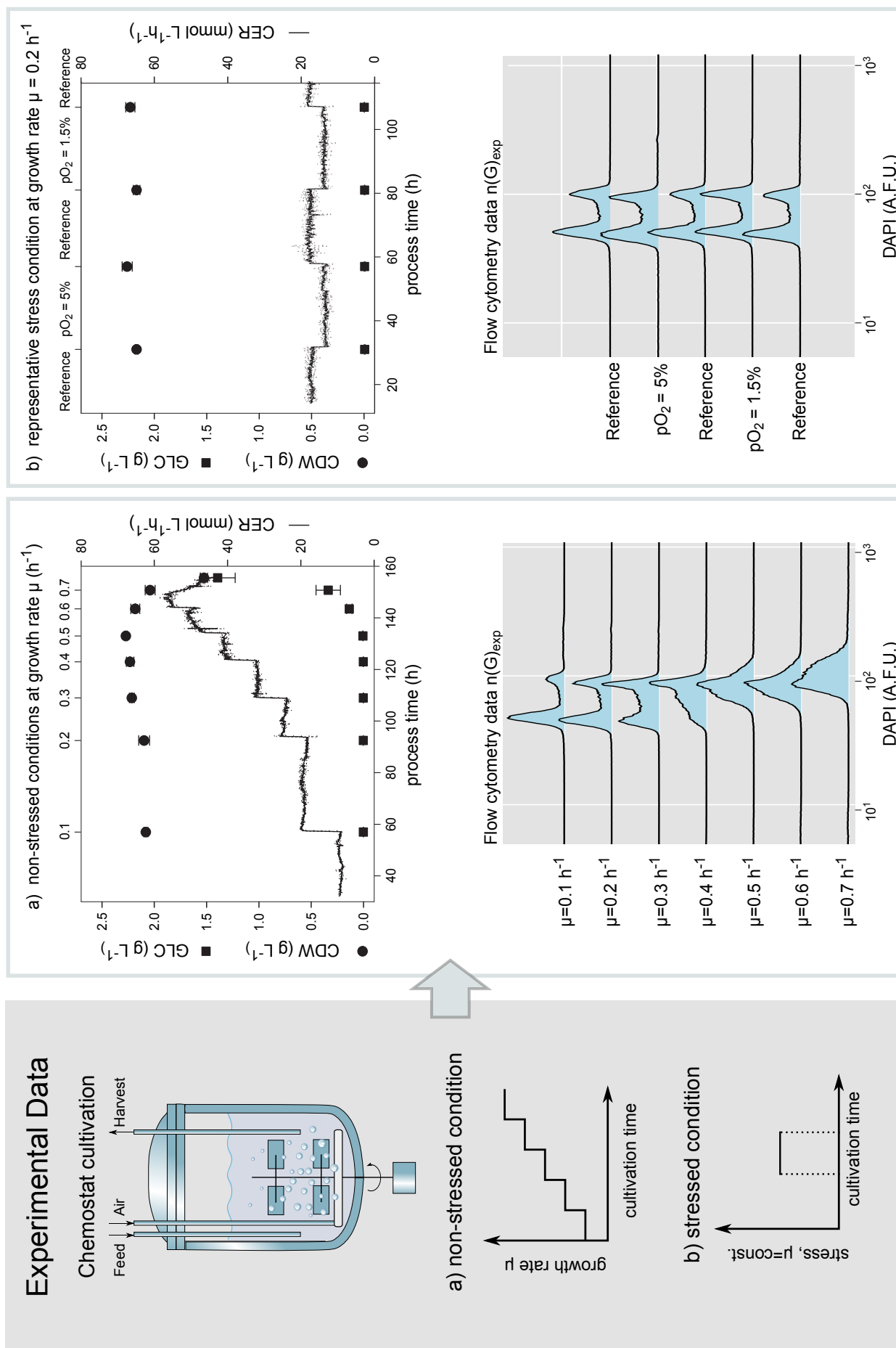


Table 7.1.: Population composition at non-stressed and stressed conditions analyzed by flow cytometry

Subpopulation	C1	C2	CX
Reference cultivation - different growth rates μ			
0.1 h ⁻¹	82.0 ± 0.3	18.0 ± 0.2	–
0.2 h ⁻¹	61.8 ± 0.9	38.3 ± 0.9	–
0.3 h ⁻¹	47.7 ± 0.5	52.3 ± 0.7	–
0.4 h ⁻¹	23.0 ± 0.9	77.0 ± 1.1	–
0.5 h ⁻¹	2.2 ± 0.8	83.1 ± 1.0	14.7 ± 0.8
0.6 h ⁻¹	2.0 ± 1.1	64.3 ± 0.9	33.7 ± 1.1
0.7 h ⁻¹	1.4 ± 0.8	16.1 ± 0.1	82.5 ± 1.0
Stress condition - constant $\mu = 0.2 \text{ h}^{-1}$			
iron -50 % w/v	60.3 ± 0.6	39.7 ± 0.3	–
pO ₂ 5 %	59.5 ± 0.7	40.5 ± 0.9	–
pO ₂ 1.5 %	52.3 ± 0.5	47.7 ± 0.5	–
decanol 5 % v/v	45.3 ± 0.8	54.7 ± 0.6	–

- (iii) subpopulation Cx with more than doubled chromosome content (found at fast growth rates), representing cells performing multifork DNA replication

Under non-stressed conditions at slow to moderate growth rates (0.1 - 0.4 h⁻¹), fractions of C1 decreased with increasing growth rate while fractions of C2 increased. At growth rates of $\mu > 0.4 \text{ h}^{-1}$, subpopulation Cx appeared. Growth rate 0.4 h⁻¹ can be defined as a threshold in *P. putida* KT2440, as cells start to uncouple DNA replication from cell division at faster growth rates. At higher growth rates than 0.4 h⁻¹, only a negligible fraction of C1, a decreasing portion of C2 and an increasing fraction of Cx were observed.

Under stressed conditions, the fraction of C1 cells decreased while more C2 cells were abundant in comparison to the reference condition at the same constant growth rate $\mu = 0.2 \text{ h}^{-1}$. This surprising phenomenon was more pronounced, the more severe the stress condition (Table 7.1).

Figure 7.1.: Overview of the experimental set-up to investigate the impact of environmental stress on population dynamics. Chemostats were carried out under non-stressed and stressed conditions. Under non-stressed conditions, all nutrients except glucose were supplied in excess and the growth rate was stepwise increased until wash-out occurred. Under stressed conditions, three different stresses were applied in a shift like manner: decreased iron availability, deprivation of oxygen and solvent exposure. Physiological data at non-stressed (a) and at a representative stress condition (oxygen deprivation, b)) are shown in the first row. Biomass density (CDW, black dots, g L⁻¹) and residual glucose concentration (GLC, black squares, g L⁻¹) were measured after five residence times of one specific dilution rate at steady state. The carbon dioxide emission rate (CER, black line, mmol L⁻¹ h⁻¹) was monitored online. The error bars and lines represent the standard deviation of biological triplicates. A summary of the flow cytometry data is given in the second row. DNA histograms (DAPI, arbitrary fluorescence units A.F.U.) are depicted for the non-stressed and the stressed experimental set-up.

7.3. Quantification of stress impact on population dynamics using mathematical modeling

Comparing the distribution of subpopulations with different DNA content under stressed and non-stressed conditions, a clear difference could be detected. Consequently, we assumed that the cell cycle was affected under stress.

A cell using binary fission for proliferation passes through three stages during its cell cycle: a stage from cell birth to initiation of replication (B phase), a replication phase (C phase) and a period between termination of replication and cell division (pre-D/D phase). Differences in fractions of subpopulations of different DNA content hint towards differences in the duration of cell cycle phases during stress exposure.

In order to answer whether and how stress imposed on bacterial growth affects the cell cycle quantitatively, a model accounting for population heterogeneity was needed to connect the fractions of subpopulations with the duration of the cell cycle phases.

7.3.1. Mathematical framework

We determined the duration of cell cycle phases by combining chemostat cultivation, flow cytometry data and mathematical simulation.

The mathematical model chosen for determining the durations of the cell cycle phases is based on the Cooper-Helmstetter cell cycle model (Cooper et al., 1968). This model allows the calculation of theoretical DNA distributions in a population having a certain generation time τ and known cell cycle durations C and D. Considering the D phase, one has to keep in mind that no distinction can be made between pre-D and D phase based on DAPI fluorescence data, because cells contain a double chromosome content in both phases. Therefore, we defined a combined parameter D' representing the time between end of replication and end of division.

The theoretical DNA distribution $n(G)$ results from linking an age distribution of a population $n(a)$ to a cellular DNA accumulation function $G(a)$.

The age distribution was implemented as a probability density function ($n(a)$):

$$n(a) = 2 \cdot \ln 2 \cdot e^{(-a \ln 2)} \quad 0 \leq a \leq 1 \quad (7.1)$$

$$\int_0^1 n(a) da = 1 \quad 0 \leq a \leq 1 \quad (7.2)$$

a is the normalized age of a cell within one generation time τ : Newly divided cells are defined to have an age $a = 0$ and cells that are dividing have the age $a = 1 = \tau$. $n(a)$ represents the

probability density of a single cell in a population to have a certain age a (Lindmo, 1982). As a consequence of binary cell division, there has to be a double amount of new born cells in comparison to dividing cells $n(0) = 2n(1)$. The amount of cells in a specific age interval can be calculated integrating the age distribution between two ages a_i and a_{ii} .

Cooper et al. (1968) assumed that the movement of the replication fork along the chromosome is constant. During one generation time, the rate of DNA synthesis was described mathematically in a step function with two discontinuities: initiation and termination of DNA replication. The specific ages a_1 and a_2 , where initiation and termination take place, were calculated as follows:

$$a_1 = (x\tau - (C + D'))/\tau \quad (7.3)$$

$$a_2 = (\tau - D')/\tau \quad (7.4)$$

Here, parameter x refers to multiples of generation times in which replication C and division D' take place.

To derive the amount of DNA per cell at a specific age $G(a)$, the division cycle was divided into three periods, defined by the ages a_1 and a_2 at the discontinuities. The chromosome content was calculated for each of these intervals as follows, considering that the chromosome content at division is the double amount of the new born cell $G(a = 1) = 2G(a = 0)$:

$$G(a) = k(F_1a + F_3) + a_1k(F_1 - F_2) + a_2k(F_2 - F_3) \quad 0 \leq a \leq a_1 \quad (7.5)$$

$$G(a) = k(F_2a + F_3) + 2a_1k(F_1 - F_2) + a_2k(F_2 - F_3) \quad a_1 \leq a \leq a_2 \quad (7.6)$$

$$G(a) = kF_3(a + 1) + 2a_1k(F_1 - F_2) + 2a_2k(F_2 - F_3) \quad a_2 \leq a \leq 1 \quad (7.7)$$

Here, F refers to the number of replication forks in the interval i . k is the constant rate of DNA synthesis per replication fork, which can be derived by $k = \tau/2C$.

The DNA distribution $n(G)$ was derived by the combination of the derivation of the theoretical chromosome content dG/da and the age distribution $n(a)$.

$$n(G)/dG = n(a)/da \quad (7.8)$$

A step-by-step illustration of the calculation of the theoretical DNA distributions can be found in Figure 7.2.

7.3.2. Implementation of the mathematical model

To correlate flow cytometry data of chemostat cultures with cell cycle dynamics, we implemented the mathematical model of Cooper and Helmstetter (1968) as described above in MATLAB®. Previously published modeling tools using the same mathematical framework were not used in this

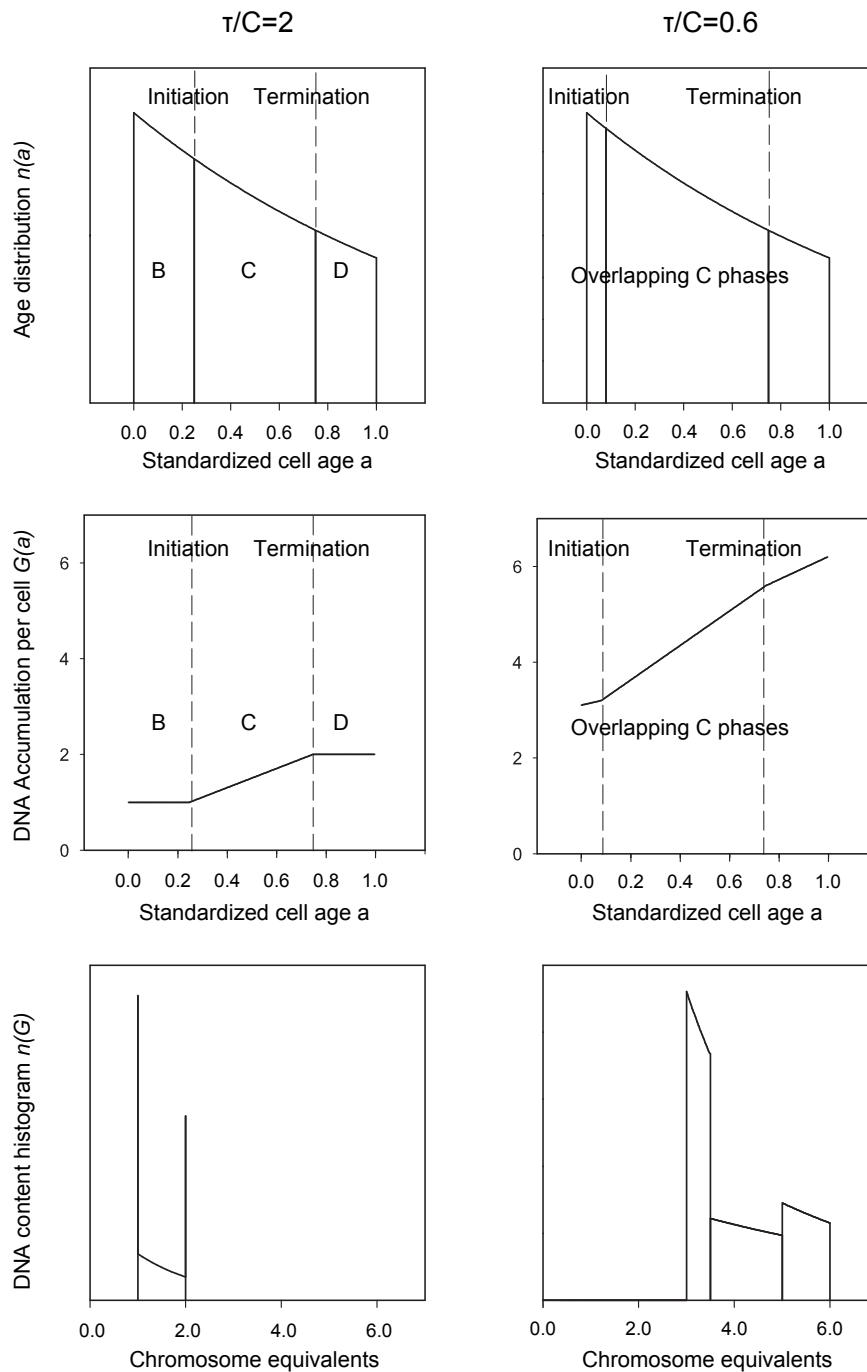


Figure 7.2.: Illustration of the calculation of DNA histograms $n(G)$. As an example, DNA histograms $n(G)$ were calculated according to Cooper and Helmstetter (1968) for a slow growing ($\tau/C = 2$) and a fast growing population ($\tau/C = 0.6$). The age distribution $n(a)$, the DNA accumulation per cell $G(a)$ and the theoretical DNA histogram $n(G)$ are illustrated for the slowly growing population (first column) and for the fast growing population (second column). Considering the mechanism of binary fission, the number of cells that has just divided doubles those that start to divide. Depending on how fast the cells are growing, the initiation and the termination of the replication shift within the timeline of a standardized cell age a . In slowly growing cells there are phases without active replication (B and D') resulting in a constant DNA content. During the replication phase the DNA content is increasing linearly with the constant rate of replication (line 2). In the case of fast growing cells, overlapping replication cycles occur, resulting in active replication throughout the age of the cell. Replicating cells of different ages can therefore have different total replication rates according to the number of replication forks at work. The portion of cells for each DNA channel in a histogram can be calculated by combining the age distribution and the DNA accumulation ($n(G)/dG = n(a)/da$, line 3).

study, because of outdated versions of computing languages (Skarstad et al., 1985), limited applicability (omitting multi-fork replication (Michelsen et al., 2003)), or because they solely allowed simulation, instead of supporting data-based parameter identification via non-linear regression (Stokke et al., 2012).

In order to simulate more ‘realistic’ DNA histograms than the theoretical histograms calculated with the mathematical framework, biological and technical measurement variations were additionally introduced, following the example of Skarstad et al. (1985):

Biological variation was simulated by slight variation of the generation time τ (coefficient of variation CV=5%). The population was artificially divided into 30 subpopulations covering the total range of variance. One resulting DNA distribution for the whole population was calculated containing all 30 simulated ‘subpopulations’.

Technical measurement variation was taken into account by assuming each DNA value in the DNA histogram to be normally distributed. The mean coefficient of variation was calculated as 5%.

Finally, the durations of the cell cycle phases were calculated via the following procedure: The generation time τ and the experimentally derived DNA histogram $n(G_{exp})$ were used as inputs for the simulation software. The cell cycle parameters C and D’ were iterated until the best fit between the experimental and theoretical DNA histogram was obtained (least-square fit Eq. 7.9), resulting in the best estimates of the cell cycle parameters in our experiments (\hat{C} and \hat{D}'). During the iteration, lower bounds of C and D’ were set to 0, while the upper bounds were set to $D' = \tau$ and $C = \tau/0.45$, respectively (Cooper et al., 1968). To evaluate our simulations we calculated the deviation s using the formula presented by Skarstad et al. (1985):

$$s = \sqrt{\frac{\sum_{i=1}^m (\sqrt{n(G_{exp})_i} - \sqrt{n(G)_i})^2}{m-1}} \quad (7.9)$$

7.3.3. Quantitative impact of stress on cell cycle phases

The simulated DNA distributions deviated only marginally from the experimentally derived DNA distributions at all growth conditions applied (Table 7.2). This accuracy is taken as evidence that the basic modeling assumptions of Cooper and Helmstetter (1968) and Skarstad (1985) can be applied for *P. putida* KT2440 as well.

Table 7.2.: Summary of the duration of cell cycle phases and goodness of fit of the simulation. Average values of calculated \hat{B} , \hat{C} and \hat{D}' phases (h) of 3 biological replicates and their standard deviation were calculated on the basis of the mathematical model of Cooper and Helmstetter (1968).

Experimental condition	\hat{C} (h)	\hat{D} (h)	\hat{B} (h)	s^1
Non-stressed cultivations - different growth rates μ				
0.1 h ⁻¹	3.48 ± 0.01	1.01 ± 0.17	2.41	0.55 ± 0.1
0.2 h ⁻¹	1.54 ± 0.04	0.94 ± 0.07	0.92	0.69 ± 0.35
0.3 h ⁻¹	1.38 ± 0.04	0.63 ± 0.06	0.29	0.49 ± 0.14
0.4 h ⁻¹	1.20 ± 0.03	0.59 ± 0.05	0	0.40 ± 0.18
0.5 h ⁻¹	1.04 ± 0.02	0.66 ± 0.01	0	0.38 ± 0.21
0.6 h ⁻¹	1.03 ± 0.02	0.57 ± 0.04	0	0.84 ± 0.26
Stress cultivations - constant $\mu = 0.2 \text{ h}^{-1}$				
iron - 50 % ²	1.07 ± 0.03	1.54 ± 0.04	0.79	0.44 ± 0.26
pO ₂ - 5 %	1.05 ± 0.03	1.50 ± 0.05	0.85	0.38 ± 0.24
pO ₂ - 1.5 %	0.94 ± 0.05	1.44 ± 0.05	1.02	0.48 ± 0.16
decanol - 5 % (v/v)	0.80 ± 0.04	1.42 ± 0.05	1.18	0.36 ± 0.15

Cell cycle analysis of non-stressed steady-state cultures at different growth rates

Under non-stressed conditions, the time needed for replication (C phase) decreased with increasing growth rate until a minimal duration of $C_{min} \approx 62 \text{ min}$ (Table 7.2). The trajectory of the C phase durations in dependency of the growth rate was very similar comparing *P. putida* KT2440 and *E. coli* (previously published data (Helmstetter, 1996)): A goodness of fit of $R^2 = 0.95$ was calculated for the merged datasets when applying the exponential model of Keasling et al. (1995), which describes the growth rate dependency of C phases for *E. coli* (Figure 7.3).

Durations of cell cycle phases have been shown to vary with growth conditions and nutrient availability, and therefore, also with growth rate (Bipatnath et al., 1998). The dependency of the C phase duration on the growth rate for *P. putida* KT2440 is in agreement with observations of Kubitschek et al. (1978) and Helmstetter et al. (1976) in *E. coli* strains. *E. coli* B/r reached a minimum C phase duration of 42 min at growth rates $\mu > 0.6 \text{ h}^{-1}$. Here, we identified a minimum C length of about 62 min for *P. putida* KT2440 at growth rates $\mu > 0.5 \text{ h}^{-1}$ (Figure 7.3). Compared to *E. coli*, *P. putida* needs to start multifork DNA replication already at slightly lower growth rates.

Surprisingly, not only the C phase duration trajectories, but also the maximum replication rates

¹The parameter s (Eq. 7.9) is the deviation of the simulated to the experimental number of cells measured by flow cytometry and presented as subpopulation distributions in DNA histograms. The formula was framed by Skarstad et al. (1985).

²The cultivation medium contained half of the iron concentration compared to the reference condition.

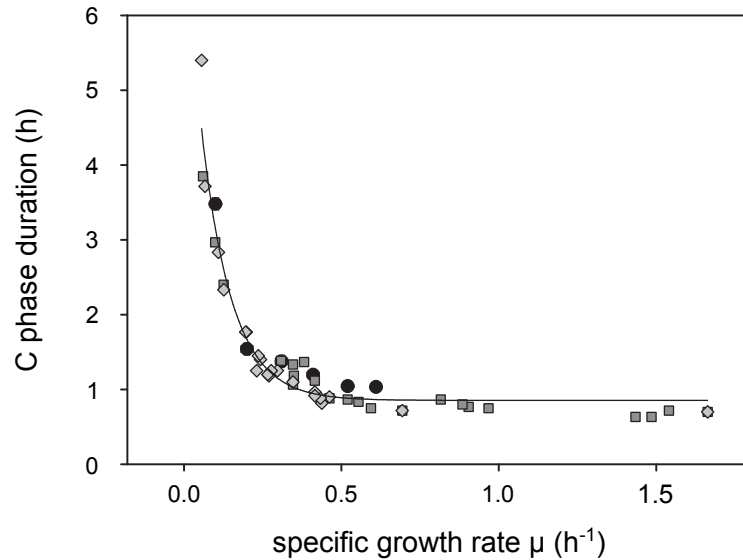


Figure 7.3.: Duration of the replication phase in dependence of the specific growth rate μ . The duration of the replication phase C was calculated according to Cooper and Helmstetter (1968) and Skarstad et al. (1985). Error bars show the standard deviation of the arithmetic mean of three biological replicates. The C phase is decreasing with increasing growth rates until a minimum duration is reached. Black dots depict the C phase durations under steady-state standard conditions of *P. putida* KT2440. Previously compiled data by Helmstetter et al. (1996) is shown as dark grey squares (*E. coli* B/r A) and light grey diamonds (*E. coli* B/r K). All data points could be reasonably well fitted ($R^2 = 0.95$) by an exponential function (Keasling et al., 1995) (black line).

of *P. putida* KT2440 and *E. coli* were highly similar. A combination of $C_{min} = 62$ min with the chromosome size of *P. putida* KT2440 (6.18 Mbp (Nelson et al., 2002)) results in a maximum replication rate of $r_c \approx 100$ kbp/min (50 kbp/min per replication fork). For *E. coli* K-12, Michelsen et al. (2003) reported a C_{min} of 46 min. Here, each replication fork travelled at a replication rate $r_c = 50$ kbp/min along the chromosome as well. The similarity between the maximal replication speeds of *P. putida* KT2440 and *E. coli* is intriguing and suggests the possibility to further exploit common properties of the replication machinery.

Cell cycle analysis of stressed steady-state cultures at constant growth rates

For the investigation of stress impact on cell cycle kinetics, chemostat cultivations were performed at a constant growth rate (0.2 h^{-1}). The duration of the cell cycle phases under reference conditions (non-stressed) were compared to different stress conditions, including reduced dissolved oxygen (pO_2) levels, exposure to the organic solvent decanol and decreased iron availability. Under all stressful conditions tested, the duration of the replication phase was shortened, while B- and D'-phases were prolonged (Figure 7.4).

A shortened C phase corresponds to an increased replication rate. At iron deprivation and low oxygen partial pressure $pO_2 = 5\%$, the replication rate rose 1.5-fold from 67 to around 99 kbp/min, equaling the maximum replication rate under non-stressed conditions at $\mu > 0.5 \text{ h}^{-1}$. Surprisingly,

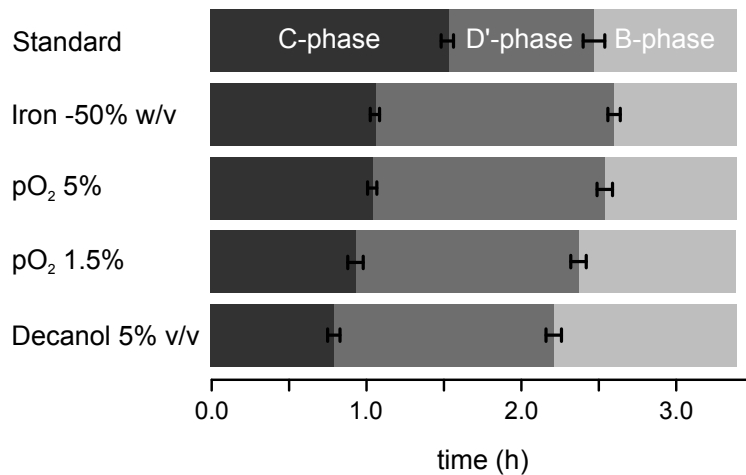


Figure 7.4.: Duration of cell cycle phases in dependence of the specific stress condition. The replication time (C phase, dark grey bar) is shown for all environmental conditions tested at a growth rate of $\mu = 0.2 \text{ h}^{-1}$. Compared to the standard conditions, the C phase decreased under all stress conditions, while D' and B phases (grey and light grey bars) increased accordingly.

when increasing the severity of the stress condition (pO₂=1.5 % and decanol exposure (5 % (v/v))), the replication rate even increased above the maximum observed at standard conditions: Lower dissolved oxygen levels of pO₂=1.5 % led to a 1.6-fold increase, while decanol exposure raised the replication rate about 1.9 fold. Notably, the generation time τ stayed constant and therefore a balanced altering of the individual contributions of B, C and D' phases was the result of the stress conditions applied.

The B phase was already described to vary in dependency of nutrient availability (Helmstetter, 1996). It was suggested, that cells need to reach a critical cell mass before entering the C phase (Donachie, 1968). This critical cell mass is either already present or rapidly reached by the cell under nutrient-rich conditions, while more time is needed in nutrient poor media. In the case of stress conditions, it is not surprising that the B phase covers part of the surplus cell cycling time.

Our results showed an extended duration between end of replication and division (pre-D / D phase). In contrast, the classical cell cycle model of Cooper and Helmstetter (1968) suggested a fixed D period. At this point, we cannot distinguish if the D phase itself is prolonged in our case, or if cells start dividing after an intermediary pre-D phase. Earlier batch cultivation studies reported a gap between end of replication and start of division in limiting conditions, leading to the introduction of the pre-D phase into the bacterial cell cycle model (Müller et al., 2003). Additionally, an enlarged, cell size generating period between end of replication and final division was found for some archaea (Lindås et al., 2013; Hjort et al., 2001) and bacteria (Robert et al., 2014). However, a prolonged D phase cannot be excluded. A delay could be a consequence of lower availability of resources which were re-distributed by the cell in favor of DNA replication or a direct mechanical interference with the divisome complex (in the case of decanol).

Table 7.3.: Differentially expressed genes under decanol stress conditions, annotated in the functional group ‘replication’. The COG database was used for functional annotation (Tatusov et al., 1997). The \log_2 fold change (FC) is the logarithmic ratio of expression of decanol condition and reference condition. Statistical significance was defined at a cutoff of the false discovery rate $FDR < 0.05$ (Benjamini et al., 1995).

Gene ID	Gene Product Name	$\log_2(\text{FC})$
PP_0979	DNA polymerase III subunit χ , HolC	1.26
PP_4141	DNA polymerase III subunit ϵ DnaQ	1.10
PP_4768	DNA polymerase III; subunit ϵ	1.02
PP_4796	DNA polymerase III subunit δ HolA	0.97
PP_4269	DNA polymerase III subunits γ and τ DnaX	0.94
PP_5310	ATP-dependent DNA helicase RecG	0.89
PP_4274	NAD-dependent DNA ligase LigA	0.67
PP_5088	Primosome assembly protein PriA	0.59

In summary, we observed a clear relationship between replication rate increase and challenging environmental conditions. In addition to previously found alterations in the cell cycle under limiting conditions, not only the time before start of replication (B phase) and the time after completion of replication until division (pre-D/D phase) increased, but also the period for replication itself was substantially altered.

Transcriptome analysis

The observation of a shortened C phase raised the question of how cells could achieve this replication speed-up. In order to get mechanistic insights, we analyzed the genome-wide expression profile of *P. putida* KT2440 via next generation sequencing and compared the mRNA levels of non-stressed cells to cells exposed to the most prominent stress condition, decanol exposure.

In total 5421 transcripts were analyzed. Among these, 540 transcripts, including 386 with annotated function, were significantly affected by decanol exposure (\log_2 fold changes (FC) > 0.58). The COG database (Tatusov et al., 1997) was used to identify 28 significantly changed genes with known or anticipated tasks in ‘replication, recombination and repair’. Thereof, 8 genes were annotated in tasks connected to ‘replication’ (Table 7.3). This group of genes showed not only a significant overexpression upon decanol exposure (average \log_2 FC 0.98), but could also be identified as significantly upregulated functional group using gene set analysis.

Especially genes that encode parts of the DNA polymerases showed elevated transcription levels (*ligA*, *holA*, *holC*, *dnaQ* and *dnaX*). Interestingly, DnaQ was shown to turn the rather slow and weakly processive polymerase III core into a fast and highly processive polymerase (Studwell et al., 1990). Furthermore, HolA increases the polymerase speed by binding the β subunit of the DNA

clamp to the polymerase core (Johnson et al., 2005), while HolC and DnaX increase the unwinding rate of the helicase DnaB (Kim et al., 1996). Altogether, the most prominent transcriptional upregulation was found for genes encoding basic enzymes that are essential for a fast, efficient and processive replication. Besides DNA replication, transcripts related to DNA repair, restart of stalled replication forks and homologous recombination were upregulated as well under decanol stress conditions (*recB*, *recD*, *recG*, *ruwC*, *mutS* and *mutL*, for details refer to section B.6, ‘Supplemental dataset’ in the appendix).

The cellular response to different types of stress is the hallmark of the cell’s strategy for survival. How organisms adjust their cell cycle dynamics to compensate for changes in environmental conditions is an important outstanding question in bacterial physiology. Our data show a clear relationship between acceleration of replication and stress exposure. We observed moderately higher expression levels of genes responsible for both key processes that determine C period duration –the velocity of the replication fork movement and the time needed for the restart of stalled replication forks (Hill et al., 2012). Consequently, we propose that the speed up of DNA replication is an actively regulated process. Previous assumptions, that (i) replication might not proceed at maximum velocity to assure stable and correct replication and that (ii) faster replication might be achieved by a higher availability of replication processivity factors (Morigen et al., 2003; Atlung et al., 2002) support our hypothesis.

Furthermore, the expression profile also hinted towards increased DNA repair and homologous recombination activity. Under stress conditions, cells might repair stress-induced errors in DNA as good as possible, while taking into account a higher frequency of recombination events. This strategy might help a population to quickly adapt to challenging stress conditions.

Summarizing, our study demonstrated that fast replication of the genetic information was of utmost priority under stress conditions. The shortened C phase was balanced by extending the duration of the remaining cell cycle phases, B and pre-D/D phases, to maintain a constant growth rate under stressed conditions.

CHAPTER 8

OPTIMIZING MICROBIAL CELL CHASSIS BY STREAMLINING THE
GENOME

This chapter contains the results and the discussion of the characterization of streamlined-genome *Pseudomonas putida* strains as microbial cell factories for heterologous protein production. Parts of this chapter have been published as ‘Genome reduction boosts heterologous gene expression in *Pseudomonas putida*’.¹

Typically, metabolic engineering approaches are applied in only a few production hosts, often *Escherichia coli* strains (Danchin, 2012; Singh, 2014). Despite their ease of genetic manipulation, these working platforms often lack desirable characteristics that are important for industrial large-scale applications (see chapter 2). In recent years, *P. putida* strains got into focus as promising alternative or extension to the bacterial working platform line-up. Especially the non-pathogenic *P. putida* strain KT2440 shows high potential, being equipped with a remarkable metabolic diversity, amenability to genetic manipulation, and stress endurance along with carrying the GRAS (generally regarded as safe) status (Kim et al., 2014; Poblete-Castro et al., 2012; Nogales et al., 2008; Nikel et al., 2014b).

The aim of a metabolic engineer is to create novel or to improve already existing production strains. On one hand, this can be done by rational pathway engineering (implementing new metabolic pathways or deleting by-product pathways). On the other hand, clearing the microbial host of all elements deemed unnecessary for cellular functions other than replication and self-maintenance might improve energy availability for production and genomic stability. Following the latter strategy of strain optimization, we analyzed kinetic and physiological parameters related to cell performance of two genome-reduced *P. putida* strains: *P. putida* EM329, lacking flagella genes,

¹Sarah Lieder[#], Pablo I. Nikel[#], Víctor de Lorenzo and Ralf Takors (2015) *Microbial Cell Factories* 14:23 [#]*Ex aequo* contribution (Appendix C)

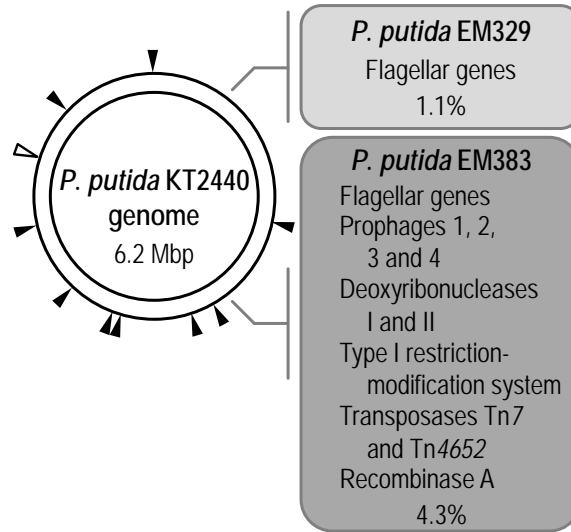


Figure 8.1.: Design of reduced-genome *P. putida* KT2440 strains. Strain EM329 lacks flagellar genes (Martínez-García et al., 2014b) while strain EM383 carries further mutations improving the strain’s cell factory characteristics (Martínez-García et al., submitted 2014). White arrowheads show the chromosomal location of the flagellar genes. Black arrowheads indicate the genes and gene clusters additionally eliminated in strain EM383. Furthermore, the extent of the deletion is noted (in percentage of the genome).

and *P. putida* EM383, carrying further mutations that were implemented to ensure genetic and physiological stability (Figure 8.1, Table C.1). Furthermore, the two multiple-deletion strains were evaluated as cell factories for heterologous protein production. We selected the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* as a model protein (Vizcaino-Caston et al., 2012) to compare production kinetics and capacities of the two manufactured strains with their parental strain.

8.1. Reaction parameters and energy profile of streamlined-genome derivatives of *P. putida* KT2440

Glucose-limited continuous cultivations at three different growth rates, $\mu = 0.1 \text{ h}^{-1}$, $\mu = 0.3 \text{ h}^{-1}$ and $\mu = 0.6 \text{ h}^{-1}$, were set up to characterize reaction parameters such as biomass yield coefficients and maintenance demands, but also to assess the energy profile of the streamlined-genome (SG) strains (for details refer to Figure C.6 in appendix C).

Biomass yield The biomass yield coefficient reflects the efficiency of substrate conversion into cell components. The yield of biomass out of glucose was calculated at steady-state conditions within the range of the investigated growth rates (Figure 8.2).

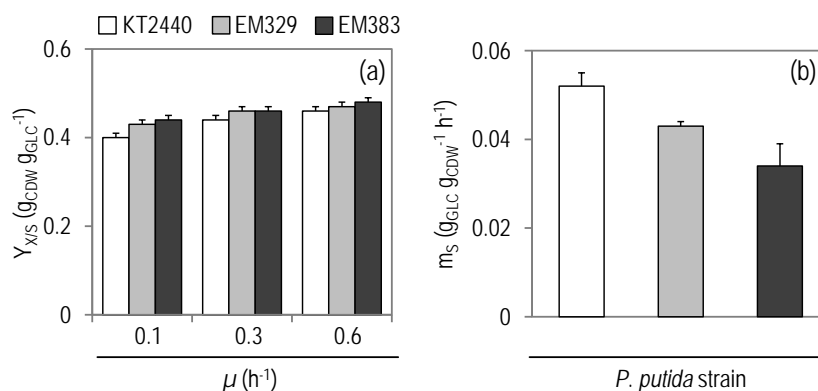


Figure 8.2.: Growth parameters of *P. putida* KT2440, EM329 and EM383 in glucose-limited chemostat cultures. (a) shows the biomass yield coefficient $Y_{X/S}$ (g_{CDW}g_{GLC}⁻¹) and (b) the maintenance coefficient m_s (g_{GLC}g_{CDW}⁻¹h⁻¹). The parameters were calculated based on three biological replicates. The bars represent the arithmetic mean of the corresponding parameter \pm standard deviations.

At all growth rates, yields were significantly higher in the derivative strains compared to the wild-type strain ($p < 0.05$). The biggest difference could be observed at the slow growth rate of $\mu = 0.1 \text{ h}^{-1}$. Here, EM383 showed a 12 % higher yield of biomass on glucose than KT2440. Notably, the differences between the two SG strains *P. putida* EM329 and EM383 were not statistically significant. Also, carbon emission rates (CER, mmol_{CO₂}L⁻¹h⁻¹, Figure C.6 in appendix C), were altered. Averaged over all growth rates, EM329 and EM383 produced 9 % and 16 % less CO₂, respectively, as compared to *P. putida* KT2440.

Maintenance coefficient The maintenance demand is an intrinsic characteristic of an organism. It reflects the amount of carbon (and ATP) needed to maintain minimal, non-growth related functions within the cell. Consequently, it is a key parameter for the choice of a microbial cell factory, since the lower the maintenance coefficient of the specific organism, the higher the carbon availability for catabolism (and biocatalysis).

The maintenance demand was calculated applying Pirt's equation (Eq. 5.24) based on the linear relationship between the specific glucose uptake rate and different growth rates (Figure 8.2 B). The wild-type *P. putida* KT2440 showed a maintenance demand of $m_s = 0.052 \pm 0.002 \text{ g}_{\text{GLC}}\text{g}_{\text{CDW}}^{-1}\text{h}^{-1}$. Notably, by-product formation could be neglected. *P. putida* does not excrete any metabolites under the conditions tested (Chavarría et al., 2013; del Castillo et al., 2007). Furthermore, all carbon balances closed within a range of $100 \pm 2 \%$, only taking biomass formation, CO₂ evolution and residual glucose concentration into account (Figure C.7 in appendix C).

Interestingly, comparing EM329 and EM383 to their parental strain KT2440, we observed 17 % and 35 % lower m_s values, respectively ($p < 0.01$). The true biomass yield coefficients, which take the maintenance demands into account, were calculated to $0.47\text{g}_{\text{CDW}}\text{g}_{\text{GLC}}^{-1}$ for strain KT2440, and $0.49\text{g}_{\text{CDW}}\text{g}_{\text{GLC}}^{-1}$ for EM329 and EM383. The differences in maintenance demands were only

significant between the derivative and the wild-type strains, but not in between the two mutant strains.

In order to estimate ATP expenses due to maintenance, we applied the following stoichiometric calculation: *P. putida* channels glucose through the Entner-Doudoroff pathway towards the tri-carboxylic acid (TCA) cycle, yielding 1 mole of ATP and 1 mole of NADH per mole of glucose consumed. In the TCA cycle, additional 4 NADH and 1 FADH per acetyl-coenzyme A are formed, which are for simplification lumped into 5 NADH during this calculation. Summarizing, 1 glucose molecule is converted into 1 ATP and ca. 11 NADH. Assuming a P/O ratio of 1.75 (Nogales et al., 2008), 21 ATP are formed during oxidative phosphorylation per glucose molecule. Consequently, m_{ATP} values ($\text{mol}_{\text{ATP}}\text{g}_{\text{CDW}}^{-1}\text{h}^{-1}$) can be calculated, resulting in 1.09 ± 0.06 for *P. putida* KT2440, and 0.91 ± 0.02 and 0.71 ± 0.05 for strains EM329 and EM383, respectively.

Our calculated maintenance demands for *P. putida* KT2440 are comparable to previously published data of van Duuren et al. (2013) under similar cultivation conditions. Maintenance coefficients of Gram-negative organisms grown in a defined glucose-containing medium varied from ca. 0.05 to $0.5 \text{ g}_{\text{GLC}}\text{g}_{\text{CDW}}^{-1}\text{h}^{-1}$ (Atkinson et al., 1967; Kooijman et al., 1992; Russell, 2007; Schulze et al., 1964). Noteworthy, the calculated maintenance demands of *P. putida* and its derivatives are rather low compared to other Gram-negative bacteria. For example, Nanchen et al. (2006) found a 28 % higher maintenance demand for the industrially well established host *E. coli* in a similar glucose limited cultivation set up. Intriguingly, the deletion of cellular components and structures that consume energy, resulted in a reduction of the maintenance demands of *P. putida*. We assume that the m_s decrease is correlated with the lack of the flagella. Energy, needed not only for the synthesis and assembly of the flagellum, but also for its operation, is saved.

Energy status Production environments challenge cell factories with an increasing energy demand. In order to assess the energetic capacity of a microbial cell, various physiological parameters, such as the ATP/ADP ratio, yield coefficients quantifying the amount of ATP or amount of total phosphorylated forms of adenine available per unit of biomass ($Y_{\text{ATP}/\text{X}}$ and $Y_{\text{AXP}/\text{X}}$, respectively), or the adenylate energy charge (AEC), can be calculated. The AEC is preferred over the ATP/ADP ratio as the parameter reflecting the energy status of the cells, as it considers the relative contribution of all three phosphorylated forms of adenine.

Both derivative strains showed a statistically significant increase in all energy parameters at all growth rates compared to KT2440 ($p < 0.01$), namely $Y_{\text{ATP}/\text{X}}$ (Figure 8.3 a), $Y_{\text{AXP}/\text{X}}$ (Figure C.3) and AEC (Figure 8.3 b). Especially under fast growth conditions, *P. putida* EM383 managed to keep a higher level of intracellular ATP and AEC in contrast to the other two strains: The difference in the ATP content between strain EM383 with respect to both EM329 and KT2440 was more than doubled (Figure 8.3). These results are fully consistent with the decreased maintenance coefficient in the SG strains explained above.

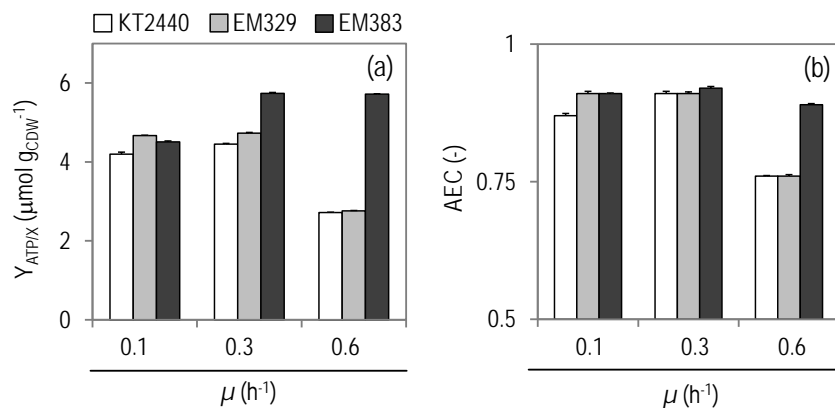


Figure 8.3.: Summary of energy parameters of *P. putida* KT2440, EM329 and EM383 in glucose-limited continuous cultures. Shown are (a) the yield of ATP on biomass ($Y_{ATP/X}$), and (b) the adenylate energy charge (AEC) of the cells at three different growth rates (μ). Bars represent the arithmetic mean of three biological replicates \pm standard deviations.

Summarizing, our characterization of the genome-reduced strains *P. putida* EM329 and EM383 stresses important physiological advantages for industrial application in terms of biomass yields, maintenance demands and energy levels over the wild-type KT2440 strain. The results suggest that saved resources to synthesize cellular components, such as flagella, lead to higher efficiency of substrate conversion into biomass, resulting in lower maintenance demands and higher energy capacity. The lower CO₂ production is an additionally interesting trait for bioprocesses depending on biomass formation. In a next step, we evaluated these potentially advantageous traits in a heterologous protein production scenario.

8.2. Heterologous protein synthesis in streamlined-genome derivatives of *P. putida* KT2440

Growth kinetics, by-product formation and recombinant protein production capacities of the SG derivatives were compared to the parental strain KT2440 in bench-top reactor batch cultivations using citrate and glucose as carbon sources. The cultivation on both, a gluconeogenic and a glycolytic carbon source, allowed a deeper analysis of the properties of these strains under different metabolic regimes.

Growth parameters In all batch cultivations, regardless of the carbon source used, the derivative SG strains reached statistically significant higher μ_{max} values than the wild-type KT2440 strain (Figure 8.4). Using glucose as sole carbon source, μ_{max} increased about 7% and 10% for EM329 and EM383, respectively (Figure 8.4 a, $p < 0.05$), while growth on citrate resulted in a 4% and 11% faster growth of EM329 and EM383 (Figure 8.4 b, $p < 0.05$).

Organic acids formation By-product secretion is an unwanted phenomenon in industrial fermentation, because carbon and cofactors, such as ATP or NADPH, are diverted from the actual production pathway. Consequently, by-product spillage reduces the production capacity of a microbial cell factory (Silva et al., 2012). As mentioned in section 8.1, *P. putida* is known for not secreting metabolites as by-products at a high concentration. This trait makes *P. putida* cultivations preferable over for example *E. coli* fermentations, where acetate is often found as unwanted by-product (Wong et al., 2008). However, using glucose as carbon source, *P. putida* oxidizes parts of the glucose to gluconate via the glucose dehydrogenase activity in the cell periplasm (del Castillo et al., 2007). From here, gluconate can leak out into the culture medium and is typically re-used as substrate at a later stage of the batch cultivation.

We investigated gluconate secretion into the cultivation medium, emphasizing on the comparison of the SG derivatives with the wild-type strain. Accumulation kinetics of gluconate during batch cultivation were very similar among all strains –*P. putida* KT2440, EM329 and EM383–, peaking in the mid-exponential growth phase. However, the gluconate found in the supernatant was metabolized completely until the end of the exponential phase, where the gluconate concentration decreased below the detection limit. Comparing the maximum accumulation of the different strains, the SG derivatives accumulated generally less gluconate than the wild-type: In the cultivation supernatant of *P. putida* KT2440, a maximum of 18.5 ± 3.1 mM (ca. 3.5 g L^{-1}) gluconate was found in contrast to 10.2 ± 1.4 and 9.3 ± 1.5 mM in EM329 and EM383 cultivations, respectively. This significant reduction of glucose oxidation to gluconate (45 % and 50 % in case of EM329 and EM383, respectively) suggests, that more carbon is readily available for catabolism.

Recombinant protein production In order to compare the capacity of the strains to produce recombinant proteins, we transformed all strains with the GFP expression plasmid pS234G. The wild-type KT2440 was additionally transformed with the empty vector pSEVA234 as a further control. Introducing the empty vector into KT2440 did not significantly influence the growth behavior (≈ 1.5 %). Consequently, a physiological effect of transforming the cells with the empty vector was neglected.

In contrast, expressing *gfp* from the plasmid pS234G led to a 6 % lower μ_{\max} of KT2440. The derivative strains showed a different growth behavior under *gfp* expression compared to their parental strain: The growth rate did not decrease significantly under heterologous protein expression conditions. Furthermore, the trend of generally higher maximum growth rates of the SG strains under normal growth conditions was also mirrored under protein expression conditions (Figure 8.4). This effect was most pronounced when growing EM329/pS234G and EM383/pS234G on citrate as carbon source (32 % faster μ_{\max} , Figure 8.4 b, $p < 0.05$).

In a next step, we monitored the GFP fluorescence during the cultivation to assess the GFP production itself. As expected, fluorescence increased exponential during exponential growth of

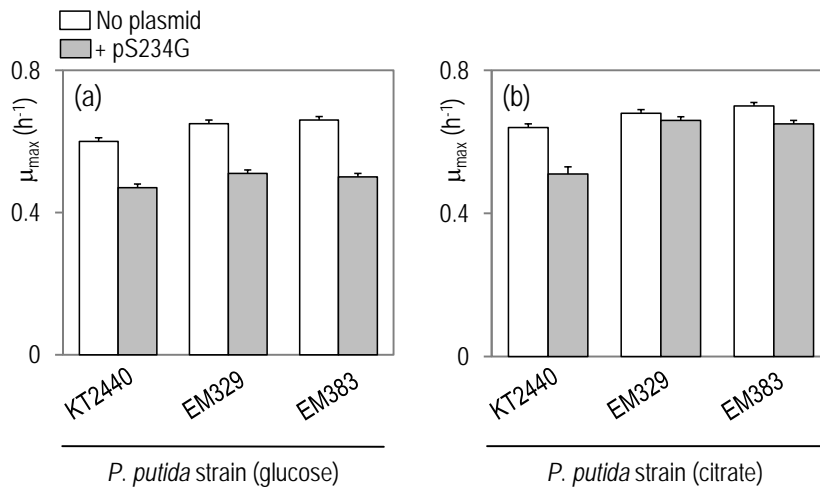


Figure 8.4.: Impact of gfp expression on the maximum specific growth rates of *P. putida* KT2440, EM329 and EM383. The maximum specific growth rate (μ_{\max}) for the different strains grown on glucose (a) and citrate (b) is shown.

the cells (Figure C.9 and Figure C.10 in appendix C). Notably, the SG strains were capable of producing significantly more GFP (Figure 8.5a). Using citrate as a carbon source, π_{\max} increased 43 % and 48 % in *P. putida* EM329/pS234G and EM383/pS234G compared to *P. putida* KT2440/pS234G ($p < 0.05$). This trend could also be found calculating the GFP production yield (Figure 8.5b): $Y_{\text{GFP}/X}$ was 18 % and 37 % higher in the derivative strains *P. putida* EM329/pS234G and EM383/pS234G, respectively, when grown on glucose. Again, this effect was strengthened when the cells were grown on citrate. Here, cells were capable of producing 20 % and 41 % more GFP per biomass as compared their parental strain.

Summarizing, the characterization of *P. putida* EM329/pS234G and EM383/pS234G showed significantly improved heterologous protein production capacities of the genome-reduced strains in comparison with the wild-type strain. In addition, the physiological advantages for industrial applications that were observed under non-producing conditions, such as faster growth, higher biomass yields and energy levels, could be maintained under production conditions.

Ideally, cell factories should behave predictably, producing the desired product –and only the desired product– with the expected yield within the expected time frame, while carrying a variety of artificial genetic constructs. The idea of complete predictability goes hand in hand with organisms containing only the minimal gene set necessary to sustain life. Many genome projects were started in the mid-1980's in order to identify this minimum number of functions. By now, we learned that the picture we had in mind was not as simple as we thought. We are still far away from complete predictable cell behavior – one of the reasons being the lack of knowledge about functionality and essentiality of a number of genes in a wide variety of environmental conditions. The minimum gene set of the environmental bacterium *P. putida* for its survival in soil is most likely not the same

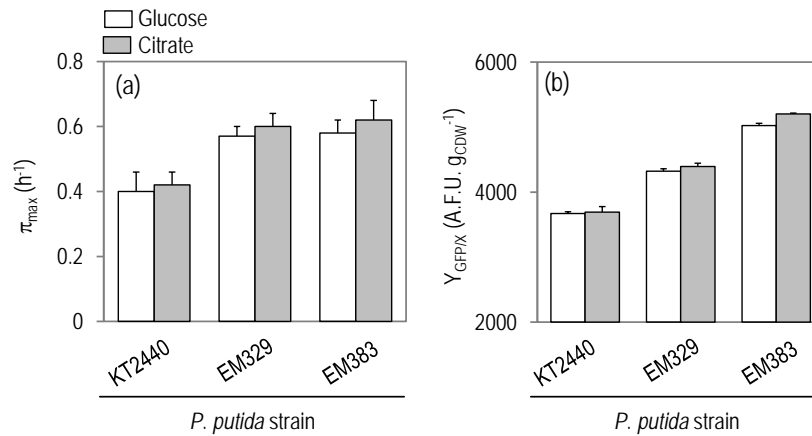


Figure 8.5.: Heterologous protein production in *P. putida* KT2440/pS234G, EM329/pS234G and EM383/pS234G bioreactor batch cultures. All strains were transformed with the *gfp* carrying expression plasmid pS234G. The GFP production was assessed during exponential growth of the culture calculating the maximum specific GFP production rate (π_{\max} , h^{-1}) (a) and the yield coefficient $Y_{\text{GFP}/X}$ (b) for the different strains grown on glucose and citrate. Bars represent the arithmetic average of three biological replicates \pm standard deviation.

minimum gene set for efficient production of heterologous proteins in an industrial cultivation setup.

The deletion of the flagellar operon in *P. putida* KT2440, which is obviously necessary for survival in its natural habitat, resulted in clear physiological advantages in our production scenario. The surplus of ATP and NADPH (Martínez-García et al., 2014b) was directly or indirectly available in the heterologous production pathway. Adding the deletion of the proviral load, which enhances stress tolerance in *P. putida* KT2440 (Martínez-García et al., 2014a), pronounced the physiological advantages for growth and production in a bioreactor even more.

The extensive ‘genomic surgery’ project of Blattner and collaborators in *E. coli* MG1655 enhanced genetic stability in their multiple deletion strains for hosting and expressing heterologous genes (Csörgo et al., 2012; Pósfai et al., 2006; Sharma et al., 2007; Umenhoffer et al., 2010). However, the significant reductions of the *E. coli* MG1655 genome size cannot overcome the retaining genomic and biochemical framework of a typical enterobacterium (Mizoguchi et al., 2007). Expression of recombinant genes, or even whole pathways, cause stress and higher ATP and/or NAD(P)H demands (Na et al., 2010; Nicolaou et al., 2010). These issues were successfully improved in the derivative *P. putida* strain, exploiting and improving the natural capabilities of the soil bacterium *P. putida*. Clearly, the side-by-side comparison of streamlined *P. putida* and *E. coli* as microbial cell factories is beyond the scope of this work. But, our results show doubtlessly the potential of derivative strains of *P. putida* as production host: The strains outcompeted the wild-type *P. putida* KT2440 strain in all biotechnologically relevant parameters that were analyzed in this study.

CHAPTER 9

CONCLUSIONS AND PERSPECTIVES

This chapter summarizes the results of this thesis on the basis of the work packages formulated in chapter 1. Furthermore, it gives conclusions to the scientific questions raised and points out future perspectives.

The cell cycle as origin of population dynamics In the first part of the thesis we investigated the role of the cell cycle as an origin of population heterogeneity in dependence on the growth rate. A number of cell performances, including product synthesis, are assumed to occur in dependency of the cell cycle phase and therefore, heterogeneity resulting from cell cycling could ultimately lead to performance loss in production processes (Jandt et al., 2014).

Proteins define the cell's functionality and the abundance of proteins reflect cell decisions. In order to detect population heterogeneity as a consequence of cell cycling, we quantified the dependency of the protein inventory of cells on different cell cycle phases under slow and fast growth conditions.

Chemostat cultivations were successfully set-up as the cultivation system of choice to ensure constant growth conditions and to clearly separate the impact of the cell cycle on population heterogeneity from any interfering, overlaying or amplifying parameter as good as possible.

Investigating subpopulations at different cell cycle stages, the parameter 'DNA content', assessed by flow cytometry, showed the strongest difference between single cells within the population and allowed to quantify the subpopulations and to directly link them to a cell cycle phase. Based on their DNA content, subpopulations that (i) just divided, but did not start replication yet, (ii) finished replication, but did not divide yet and (iii) carried out multifork replication were sorted via fluorescence activated cell sorting and the 'sub-proteome' of the subpopulations was assessed using label free mass spectrometry (UfZ Leipzig).

Summarizing, the protein inventory of a cell was highly similar and therefore independent of the cell cycle phase, regardless of the growth rate investigated. The hypothesis that cells in different cell cycle stages specialize into e.g. carbon storage or protein production/growth, especially in B- and pre-D phases, could not be supported. This result is remarkable, as it gives rise to the assumption that the cell cycle itself has a minor impact on population heterogeneity on the level of proteome under the conditions tested.

Comparing the effect of the cell cycle phase and the growth rate on the cellular protein composition, the growth rate played a superior role in determining the functional diversity of cells within a population. Interestingly, no higher abundance of proteins related to energy or carbon metabolism could be detected in dependence on the growth rate. Therefore, we assume that higher specific glucose uptake rates at fast growth were only accompanied by higher absolute protein quantity, resulting in no change of the relative quantity of proteins in these metabolic pathways.

As an extension of this study, it would be interesting to investigate the subpopulation proteome of a *P. putida* KT2440 strain that produces the green fluorescent protein GFP heterologously. In order to prevent biased heterogeneity information due to variability in plasmid copy numbers, it would be important to construct a strain with a genomic *gfp* insertion. Applying the same experimental workflow as presented here, the cell sorting strategy based on ‘DNA content’ could be extended by ‘GFP fluorescence’ and therefore, would give another layer of information on cell physiology under production conditions in direct comparison with the results obtained in this study.

The environmental condition as origin of population dynamics Stress-shift chemostats were combined with mathematical modeling to investigate the impact of industrial relevant stress conditions on population heterogeneity. Oxygen deprivation, decreased iron availability and solvent exposure were chosen as representative stress conditions occurring in industrial cultivations. The stress-shift chemostat set up at a constant growth rate was developed and tested for the different stresses and successfully carried out in biological triplicates with a variation of less than 7%.

We quantified the subpopulations that arose under stress conditions via flow cytometry. The distribution of cells with different DNA content was clearly altered between the non-stressed reference and the stress condition. To be able to translate this observation into a quantifiable biological impact of stress on the population, a mathematical framework that correlates DNA content distributions to cell cycle phase durations was chosen.

The mathematical framework (Cooper et al., 1968; Skarstad et al., 1985) was implemented into MATLAB and the duration of the cell cycle phases C, D and B was successfully calculated. The simulated best-fit DNA histograms were highly similar to the experimentally derived DNA histograms, confirming that the mathematical model is also valid to use for *P. putida* KT2440 research.

Furthermore, the standard deviation of the cell cycle phase duration in biological triplicates was less than 5 %, showing that the combination of chemostat cultivations, flow cytometry and modeling is a reliable and reproducible tool for the investigation of cell cycle phases.

Under non-stressed conditions, we found not only similar growth rate dependencies of the cell cycle phases C and D comparing *P. putida* KT2440 with previously published data of *E. coli* B/r strains (Helmstetter, 1996), but also highly similar maximum replication rates. We hypothesize that these bacteria might share common principles of the replication machinery which ultimately lead to the observation of similar maximum replication rates.

Under stress conditions, the cells altered their cell cycle substantially. Cells spent more time in cell phases between division and start of replication (B phase) and between completion of replication until division (pre-D / D phase), while the duration of replication itself (C phase) was shortened. Consequently, the replication rate was accelerated. This phenomenon was enforced with the severity of the stress imposed (up to 1.9 fold).

In order to shed light on the mechanism of replication speed up, we compared RNA levels of genes annotated in the functional group ‘replication, recombination and repair’ at standard conditions with decanol stress conditions (‘whole transcriptome shotgun sequencing’). Genes associated with DNA replication and repair were significantly upregulated under stress conditions (average \log_2 FC 0.98). Therefore, increased resources of replication machinery related proteins could be a reason for the speed up of replication.

Regardless of the biological implications or exact mechanistic understanding, we found that fast replication of the genetic information is of utmost priority under stress conditions. We hypothesize that the higher expression of genes involved in DNA replication hints towards an actively regulated acceleration of DNA replication under the environmental stress conditions tested. The biological reason behind replication speed up as a survival response of *P. putida* KT2440 remains unclear. Cells may try to circumvent environmental stress by repairing stress-induced errors in DNA as good as possible and allowing recombination events to happen at a higher frequency, which may result in a quicker evolutionary adaptation capacity of the population. However, a balanced altering of not only B and pre-D phase, but also of the replication phase C itself, is the basis for a cellular strategy to cope with stress and maintaining a constant growth rate.

We showed that a combination of carefully designed experiments and mathematical modeling gives important mechanistic insights into the origin of population heterogeneity. This work is a valuable contribution to model and predict realistic population behavior: Future implementation of this mechanistic model into structured and segregated approaches, such as population balance equation models (refer to chapter 4), will certainly shed light on the dynamic emergence of subpopulations under industrially relevant cultivation conditions.

With advancing single cell analytics, especially in combination with systems level 'omics technologies, the research community will get step by step closer to decipher population heterogeneity. The transcriptome analysis carried out in chapter 7 allowed to gain insights into transcriptional upregulation of the replication machinery under stressful conditions. Considering the shortened replication time under stress, it would be intriguing to see, if overexpression of the respective genes could lead to a shortened replication time also under non-stressed conditions, therefore, possibly leading to shorter generation times of the organism –an interesting trait for biotechnological application.

Combined assessment of transcriptome data and the proteome of subpopulations –allowing implications on the actual physiological status of the cells– will lead to an even more comprehensive understanding of the physiology of *Pseudomonas* and its population behavior. Thereby, not only information on how to incorporate heterogeneity into the design and optimization of robust biotechnological processes, but also systems-guided optimization strategies for a robust microbial cell factory itself will be gained.

Optimizing *P. putida* as microbial cell factory by streamlining the genome We set up controlled batch and chemostat cultivations in order to evaluate the worth of two genome reduced *P. putida* strains as microbial cell factories in comparison to their parental strain KT2440. The first strain, *P. putida* EM329, lacked genes of the flagella machinery (Martínez-García et al., 2014b), while the second strain, *P. putida* EM383, carried further mutations regarding its prophages (Martínez-García et al., submitted 2014). Biological triplicate cultivations for every strain under each cultivation condition were successfully carried out with an averaged biological variation of less than 6 %.

The streamlined-genome (SG) strains outcompeted the parental strain in all industrially relevant physiological parameters that were investigated, *P. putida* EM383 being superior or equal to *P. putida* EM329. Targeted genome reduction affected the physiology of the strain in favor of 'most wanted' traits in industry: the maintenance demand decreased 35 %, accompanied by higher conversion of substrate into biomass, especially reflected at low growth rates (12 % increase of $Y_{X/S}$ in EM383). Consequently, also the CO_2 production rate decreased (16 %, averaged over all growth rates, in EM383) –a valuable side effect for industrial application. Furthermore, the energy capacity of the SG strains improved: The energy charge and the ATP content per biomass $Y_{ATP/X}$ increased at all growth rates. Outstandingly and in contrast to the parental strain, EM383 managed to keep its energy level high at fast growth rates, resulting at a doubled $Y_{ATP/X}$ under these growth conditions compared to KT2440.

Secondly, we assessed the potential of streamlining the genome as a strategy to optimize the heterologous protein production capacity of the cell factory. All strains were transformed with the *gfp* expression plasmid pS234G. KT2440 was additionally transformed with the empty vector pSEVA234. While the introduction of the empty vector resulted in no significant changes of

growth physiology, the expression of *gfp* caused a significant decrease of maximal growth rate μ_{\max} in the wild-type strain. On the contrary, the SG strains were less affected by the burden of heterologous protein production and even a maximum increase of 41 % of GFP per biomass could be achieved in EM383. Obviously, streamlining the genome of *P. putida* KT2440 and deleting cellular components, such as flagella, resulted in physiological advantages for industrial application purposes. Saved resources from the production, assembly and motion of the flagella resulted in a direct surplus of ATP and NADPH (Martínez-García et al., 2014b). We assume that the saved resources led to a higher substrate-to-biomass conversion efficiency and could have been channeled into heterologous protein production.

We showed that targeted streamlining of the genome can be successfully used to optimize the energy household and production capacity of microbial cell factories. The derivative strains highlighted the potential of *P. putida* strains as production hosts. In general, we promote the non-pathogenic *P. putida* KT2440 as an optimal choice as a production platform. The strain unites important characteristics for biotechnological application: a high level of stress robustness, metabolic diversity, a relative ease of genetic manipulation and a GRAS status (generally regarded as safe). While our results promote the streamlined genome *P. putida* EM329 and EM383 strains as individually sound production hosts, they also constitute a promising basis for further insights on the minimal gene set needed to maintain cell fitness and robustness and will enhance the art of tailoring production hosts for industrial needs.

Finally, a combination of targeted genome reduction and classical process parameter optimization will certainly enhance the overall production performance of *P. putida* strains in diverse biotechnological applications.

AUTHOR CONTRIBUTIONS

This chapter sums up my individual contributions to the manuscripts published in international peer-reviewed scientific journals. The content of the manuscripts is attached in the appendix.

Manuscript I

Sarah Lieder, Michael Jahn, Jana Seifert, Martin von Bergen, Susann Müller and Ralf Takors (2014) Subpopulation-proteomics reveal growth rate, but not cell cycling, as a major impact on protein composition in *Pseudomonas putida* KT2440. *AMB Express* 4:71

Sarah Lieder designed the study, carried out the chemostat cultivations, analyzed the flow cytometry and proteome datasets and drafted the manuscript.

Manuscript II

Sarah Lieder, Michael Jahn, Joachim Koepff, Susann Müller and Ralf Takors (2016) Environmental stress speeds up DNA replication in *Pseudomonas putida* in chemostat cultivations. *Biotechnology Journal* 11(1):155-63

Sarah Lieder designed the study, carried out the chemostat cultivations, implemented the mathematical model, analyzed the flow cytometry datasets, carried out the transcriptome data analysis and drafted the manuscript.

Manuscript III

Sarah Lieder[#], Pablo I. Nickel[#], Víctor de Lorenzo and Ralf Takors (2015) Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. *Microbial Cell Factories* *Microbial Cell Factories* 14:23 [#]*Ex aequo* contribution

Sarah Lieder designed the study, carried out the batch and chemostat cultivations, characterized the enhanced process parameters and energy profile of streamlined-genome derivatives of *P. putida* KT2440 in continuous cultures, evaluated the derivative strains as hosts for heterologous protein synthesis in batch cultivation and drafted the manuscript.

ACKNOWLEDGEMENTS

I am grateful to many people for their support and encouragement which were invaluable to reach this point of formulating the acknowledgements for my PhD thesis.

Firstly, I would like to thank Prof. Ralf Takors for giving me the opportunity to be a part of the inspiring European project ‘Pseudomonas 2.0’ and his supervision and mentoring throughout my doctoral studies. The always open door for motivational and advisory words was very appreciated. I am grateful to Prof. Han de Winde for the scientific discussions and advice throughout the project meetings and for being part of my thesis committee. Additional thanks go to the head of the thesis committee, Prof. Bernhard Hauer, and to the co-referees Prof. Thomas Hirth, apl. Prof. Dieter Jendrossek, apl. Prof. Jürgen Pleiss and Dr. Martin Siemann-Herzberg for their time and interest in my doctoral thesis.

I am indebted to the ERA-IB ‘Pseudomonas 2.0’ consortium. The fruitful discussions at wonderful locations led to valuable scientific input and inspiration and were highlights during my time as a PhD student. I am grateful to Prof. Susann Müller and Michael Jahn at the UfZ Leipzig, without whose collaboration this thesis would not have been possible. Susann introduced me to the power of flow cytometry analysis and enlightened my fascination for it. I am more than grateful for Michael’s patience, dedication and time while measuring my samples at the flow cytometer, his help and guidance during data analysis and writing our shared manuscripts. Working together has been a pleasure.

A special thanks goes to Pablo Nikel (CNB-CSIC, Madrid). The joy and enthusiasm he has for his research was contagious and motivational for me. Even more, I enjoyed our stays after the consortium meetings, the many coffees we shared and the scientific and non-scientific discussions that led to fruitful collaboration and friendship. I appreciate all his valuable insights and contributions of time while proof-reading this thesis.

The members of the Institute of Biochemical Engineering have contributed immensely to my personal and professional time in Stuttgart. I would like to express my thanks to Martin Siemann-Herzberg and Bastian Blombach, whose doors were always open to discuss ideas, thoughts or doubts. The group has been a source of friendship as well as good advice and collaboration and made me enjoy most days at work. I would like to thank Alex, Michael and Salah, who always helped with little or big problems during fermentations, even if it was Friday afternoon. Special thanks go to Andreas, whose enthusiasm and problem solving skills saved more than one of my cultivations and with whom I had the honor of spending time in the mountains and snow, philosophizing about outdoor equipment and life. Mira's help regarding the HPLC analysis and Martina's support and knowledge regarding shipping procedures, chemical waste deposition and lab safety, but even more the laughter and good time we shared was greatly appreciated.

It has been a pleasure to share the –sometimes hard– path of pursuing a PhD with Attila, Jens, Joana, Maria, Michael and Tobias. Tobias generously shared his cultivation skills in the very beginning of my PhD with me, which saved me a lot of time and nerves. Sharing the office with Jens and Joana contributed immensely to the joy of the every day work. Their useful and not so useful, adequate and inadequate comments and jokes but also their insightful questions, valuable ideas regarding fermentation and transcriptome analysis helped me undoubtedly to finish this work. I enjoyed spending time with you –sometimes literally night and day, while running fermentations or fighting with programming problems in Matlab or R. An additional thanks goes to Branka, who helped me keeping my office space clean, my plants watered, always found time to listen and even provided a warm welcome at her home during our holidays.

Furthermore, I am grateful to my students Christiane, Felix, Joachim, Mario, Renana and Telma for their enthusiasm and ideas and for firing always the right questions towards me. I learned a lot from you, especially in terms of self- and work organization.

Our group's administrative assistant and good soul Silke Reu kept me organized, reminded me about deadlines and was always ready to help. I am indebted to her personal commitment and patience regarding my –repetitive– questions about travel expense reporting, receiving private packages at the office and finally, convincing me about the charm of the 'Swabians' (which undoubtedly has been a challenge considering my stubborn northern character).

Last but by no means least, I thank my family for their encouragement and support in my life. I am happy to have my loving, encouraging and patient partner Bouke at my side, whose faithful support during the final stages of this PhD is so appreciated. Finally, I am indebted to the baby bump –now Lotte– for the distracting, but wonderful moments during data analysis, manuscript and thesis writing, who at a later stage taught me working hours and multitasking that I never imagined before.

Thank you.

CURRICULUM VITAE

Personal Details

Date and place of birth 24.11.1985, Rotenburg (Wümme), Germany

Education

- 06/2011 - **PhD student** at the Institute of Biochemical Engineering, University of Stuttgart
'Deciphering Population Dynamics as a Key for Process Optimization'
under supervision of Prof. Dr. Ralf Takors
- 10/2008 - 09/2010 **Studies in Biotechnology (Master of Science)**
Technical University of Braunschweig, Germany
Field of Concentration: 'Biochemical Engineering'
Final Mark 1.2 (1: very good - 5: insufficient)
- 03/2010 - 09/2010 Master Thesis at Novozymes A/S, Kalundborg, DK
'Fermentation physiology of *Bacillus licheniformis*: A platform for extracellular protein production'
- 11/2009 - 03/2010 Internship at Novozymes A/S, Kalundborg, DK, Fermentation Optimization
- 10/2005 - 10/2008 **Studies in Biotechnology (Bachelor of Science)**
Technical University of Braunschweig, Germany
Field of Concentration: 'Biochemical Engineering'
Final Mark 1.6 (1: very good - 5: insufficient)
- 03/2008 - 06/2008 Bachelor Thesis at the Center for Microbial Biotechnology, DTU, Denmark
'Investigation of the Role of Tyrosine Phosphatases in *Streptomyces coelicolor*'

Work Experience

- 10/2015 - Scientist at Zymergen, Inc. (San Francisco, CA, USA)
- 10/2010 - 05/2011 Researcher in the cooperation project ‘Computational analysis of genetic interactions and deletion case-studies towards improved genotype-phenotype predictions’ between the groups ‘Architecture and Regulation of Metabolic Networks’ (Dr. Kiran Patil, EMBL Heidelberg) and ‘BioProcess Systems Engineering’ (Assist. Prof. Isabel Rocha, Universidade do Minho, Portugal)
- 2007 - 2010 Institute of Machine Tools and Production Technology, TU Braunschweig
Research and Teaching Assistant
- 2006 - 2007 Institute of Biotechnology, TU Braunschweig
Teaching Assistant (Mathematics Tutorial)

Awards

- 03/2010 -09/2010 DAAD Scholarship for short term research projects (German Academic Exchange Service)
- 03/2008 - 06/2008 ERASMUS Freemover scholarship
- 10/2007 - 10/2008 Technical University of Braunschweig Scholarship ‘Best of class’

Additional Qualifications

- Languages English (fluent, scientific writing skills)
 Spanish (conversational), Danish (conversational)
 German (native language)
- Computing MS-Office (very good knowledge)
 LaTeX (very good knowledge)
 R (good knowledge)
 MATLAB (basic knowledge) and LabVIEW (basic knowledge)
- Personal Interests My classic VW Beetle, playing the piano, water sports

San Francisco, August 20, 2016

REFERENCES

- Ackermann JU, Müller S, Lösche A, Bley T, Babel W (1995).** *Methylobacterium rhodesianum* cells tend to double the DNA content under growth limitations and accumulate PHB. *Journal of Biotechnology* 39.1:9–20 (cit. on pp. 14, 37, 101).
- Adams DW, Errington J (2009).** Bacterial cell division: assembly, maintenance and disassembly of the Z ring. *Nature Reviews Microbiology* 7.9:642–653 (cit. on pp. 42, 107, 125).
- Adams J (2008).** Transcriptome: connecting the genome to gene function. *Nature Education* 1:195 (cit. on p. 17).
- Aldor IS, Keasling JD (2003).** Process design for microbial plastic factories: metabolic engineering of polyhydroxyalkanoates. *Current Opinion in Biotechnology* 14.5:475–483 (cit. on p. 110).
- Almquist J, Cvijovic M, Hatzimanikatis V, Nielsen J, Jirstrand M (2014).** Kinetic models in industrial biotechnology – Improving cell factory performance. *Metabolic Engineering* 24:38–60 (cit. on p. 132).
- Anders S (2010).** HTSeq: Analysing high-throughput sequencing data with Python. URL: <http://www.huber.embl.de/users/anders/HTSeq/doc/overview.html> (cit. on p. 118).
- Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD (2013).** Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature Protocols* 8.9:1765–1786 (cit. on pp. 30, 118).
- Andrews S (2010).** FASTQC. A quality control tool for high throughput sequence data. URL: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (cit. on p. 118).
- Atkinson DE, Walton GM (1967).** Adenosine triphosphate conservation in metabolic regulation: rat liver citrate cleavage enzyme. *Journal of Biological Chemistry* 242.13:3239–3241 (cit. on pp. 64, 102, 103, 143).
- Atlung T, Hansen FG (2002).** Effect of different concentrations of H-NS protein on chromosome replication and the cell cycle in *Escherichia coli*. *Journal of Bacteriology* 184.7:1843–1850 (cit. on pp. 59, 126).

- Avery SV (2006)**. Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews Microbiology* 4.8:577–587 (cit. on pp. 2, 3, 11, 12, 37, 100).
- Bagdasarian M, Lurz R, Rückert B, Franklin FCH, Bagdasarian MM, Frey J, Timmis KN (1981)**. Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF 1010-derived vectors, and a host-vector system for gene cloning in *Pseudomonas*. *Gene* 16:237–247 (cit. on pp. 8, 25, 101, 136).
- Bailey JE (1998)**. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnology Progress* 14.1:8–20 (cit. on pp. 22, 24).
- Barry WT, Nobel AB, Wright FA (2005)**. Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinformatics* 21.9:1943–1949 (cit. on p. 31).
- Benjamini Y, Hochberg Y (1995)**. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*:289–300 (cit. on pp. 30, 58, 118, 124).
- Bipatnath M, Dennis PP, Bremer H (1998)**. Initiation and velocity of chromosome replication in *Escherichia coli* B/r and K-12. *Journal of Bacteriology* 180.2:265–273 (cit. on p. 55).
- Blank LM, Ionidis G, Ebert BE, Bühler B, Schmid A (2008)**. Metabolic response of *Pseudomonas putida* during redox biocatalysis in the presence of a second octanol phase. *FEBS Journal* 275:5173–5190 (cit. on pp. 7, 101).
- Bley T (1990)**. State-structure models – A base for efficient control of fermentation processes. *Biotechnology Advances* 8.1:233–259 (cit. on pp. 3, 14, 101).
- Bley T (2011)**. From single cells to microbial population dynamics: modelling in biotechnology based on measurements of individual cells. *High Resolution Microbial Single Cell Analytics*. Springer (cit. on p. 22).
- Brass J, Hoeks F, Rohner M (1997)**. Application of modelling techniques for the improvement of industrial bioprocesses. *Journal of Biotechnology* 59.1:63–72 (cit. on p. 24).
- Brehm-Stecher BF, Johnson EA (2004)**. Single-cell microbiology: tools, technologies, and applications. *Microbiology and Molecular Biology Reviews* 68.3:538–559 (cit. on pp. 19, 100).
- Bremer H, Dennis PP (2004)**. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology*. Ed. by Neidhardt FC. Vol. 2. ASM press Washington, DC (cit. on pp. 45, 110).
- Buchholz A, Takors R, Wandrey C (2001)**. Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectrometric techniques. *Analytical biochemistry* 295.2:129–137 (cit. on p. 103).
- Bull AT (2010)**. The renaissance of continuous culture in the post-genomics age. *Journal of industrial microbiology and biotechnology* 37.10:993–1021 (cit. on pp. 15, 127).

- Carlquist M, Fernandes RL, Helmark S, Heins AL, Lundin L, Sörensen SJ, Gernaey KV, Lantz AE (2012)**. Physiological heterogeneities in microbial populations and implications for physical stress tolerance. *Microbial Cell Factories* 11.1:1–13 (cit. on pp. 4, 14, 47, 48, 101).
- Casti J (1997)**. *Reality Rules, The Fundamentals*. Vol. 1. A Wiley-Interscience Production. John Wiley and Sons, Inc. (cit. on p. 24).
- Chavarría M, Nikel PI, Pérez-Pantoja D, de Lorenzo V (2013)**. The Entner-Doudoroff pathway empowers *Pseudomonas putida* KT2440 with a high tolerance to oxidative stress. *Environmental Microbiology* 15.6:1772–1785 (cit. on pp. 7, 63, 143).
- Chen X, Zhou L, Tian K, Kumar A, Singh S, Prior BA, Wang Z (2013)**. Metabolic engineering of *Escherichia coli*: A sustainable industrial platform for bio-based chemical production. *Biotechnology Advances* 31.8:1200–1223 (cit. on pp. 5, 133).
- Chevance FF, Hughes KT (2008)**. Coordinating assembly of a bacterial macromolecular machine. *Nature Reviews Microbiology* 6.6:455–465 (cit. on p. 152).
- Chien AC, Hill NS, Levin PA (2012)**. Cell size control in bacteria. *Current Biology* 22.9:R340–R349 (cit. on pp. 111, 114).
- Choi KH, Kumar A, Schweizer HP (2006)**. A 10-min method for preparation of highly electrocompetent *Pseudomonas aeruginosa* cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. *Journal of Microbiological Methods* 64.3:391–397 (cit. on pp. 26, 138).
- Cooper S (1991)**. *Bacterial growth and division: biochemistry and regulation of prokaryotic and eukaryotic division cycles*. Academic Press (cit. on pp. 13, 14, 100, 106, 115, 125).
- Cooper S, Helmstetter C (1968)**. Chromosome replication and the division cycle of *Escherichia coli* B/r. *Journal of Molecular Biology* 31:519–540 (cit. on pp. 13, 51–57, 70, 114, 115, 118, 121, 122, 124, 125, 128–131).
- Cox J, Mann M (2008)**. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* 26.12:1367–1372 (cit. on pp. 29, 104).
- Cserjan-Puschmann M, Kramer W, Duerschmid E, Striedner G, Bayer K (1999)**. Metabolic approaches for the optimisation of recombinant fermentation processes. *Applied Microbiology and Biotechnology* 53.1:43–50 (cit. on p. 103).
- Csete ME, Doyle JC (2002)**. Reverse engineering of biological complexity. *Science* 295.5560:1664–1669 (cit. on p. 17).
- Csörgo B, Fehér T, Tímár E, Blattner FR, Pósfai G (2012)**. Low-mutation-rate, reduced-genome *Escherichia coli*: an improved host for faithful maintenance of engineered genetic constructs. *Microbial Cell Factories* 11.11 (cit. on pp. 68, 152).
- Danchin A (2012)**. Scaling up synthetic biology: Do not forget the chassis. *FEBS Letters* 586.15:2129–2137 (cit. on pp. 5, 61, 132).

- De Lorenzo V (2014)**. From the selfish gene to selfish metabolism: revisiting the central dogma. *BioEssays* 36.3:226–235 (cit. on p. 7).
- De Marco A (2013)**. Recombinant polypeptide production in *E. coli*: towards a rational approach to improve the yields of functional proteins. *Microbial cell factories* 12.1:101 (cit. on p. 151).
- Del Castillo T, Ramos JL, Rodríguez-Herva JJ, Fuhrer T, Sauer U, Duque E (2007)**. Convergent peripheral pathways catalyze initial glucose catabolism in *Pseudomonas putida*: genomic and flux analysis. *Journal of Bacteriology* 189.14:5142–5152 (cit. on pp. 63, 66, 143, 151).
- Delvigne F, Goffin P (2014)**. Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations. *Biotechnology Journal* 9.1:61–72 (cit. on p. 100).
- Dhar N, McKinney JD (2007)**. Microbial phenotypic heterogeneity and antibiotic tolerance. *Current Opinion in Microbiology* 10.1:30–38 (cit. on p. 12).
- Díaz Ricci JC, Hernández ME (2000)**. Plasmid effects on *Escherichia coli* metabolism. *Critical Reviews in Biotechnology* 20.2:79–108 (cit. on p. 148).
- Díaz M, Herrero M, García LA, Quirós C (2010)**. Application of flow cytometry to industrial microbial bioprocesses. *Biochemical Engineering Journal* 48.3:385–407 (cit. on p. 3).
- Dittrich W, Göhde W (1969)**. Impulsfluorometrie bei Einzelzellen in Suspensionen. *Zeitschrift für Naturforschung* 24b:221–228 (cit. on p. 18).
- Donachie W (1968)**. Relationship between cell size and time of initiation of DNA replication. *Nature* 219:1077–1079 (cit. on pp. 13, 57, 110, 115).
- Dos Santos V, Heim S, Moore E, Strätz M, Timmis K (2004)**. Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. *Environmental Microbiology* 6:1264–1286 (cit. on pp. 8, 101).
- Dunn N, Gunsalus I (1973)**. Transmissible plasmid coding early enzymes of naphthalene oxidation in *Pseudomonas putida*. *Journal of Bacteriology* 114.3:974–979 (cit. on p. 8).
- Ellis B, Gentleman R, Hahne F, Meur NL, Sarkar D (2013)**. flowViz: Visualization for flow cytometry. *R package version 0.2.1* (cit. on p. 29).
- Ellis B, Haaland P, Hahne F, Le Meur N, Gopalakrishnan N, Spidlen J (2014)**. flowCore: Basic structures for flow cytometry data. *R package version 1.11.20* (cit. on p. 29).
- Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger E, Schweder T, Hamer G, O’Beirne D (2001)**. Physiological responses to mixing in large scale bioreactors. *Journal of Biotechnology* 85.2:175–185 (cit. on pp. 3, 4, 14, 47).
- Ferenci T (2006)**. A cultural divide on the use of chemostats. *Microbiology* 152.5:1247–1248 (cit. on pp. 16, 127).
- Foley P, Shuler M (2010)**. Considerations for the design and construction of a synthetic platform cell for biotechnological applications. *Biotechnology and Bioengineering* 105.1:26–36 (cit. on pp. 5, 132).

- François J, Parrou JL (2001)**. Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews* 25.1:125–145 (cit. on p. 110).
- Fredrickson AG (2003)**. Population balance equations for cell and microbial cultures revisited. *AIChE Journal* 49.4:1050–1059 (cit. on p. 22).
- Fredrickson A, Ramkrishna D, Tsuchiya H (1967)**. Statistics and dynamics of procaryotic cell populations. *Mathematical Biosciences* 1.3:327–374 (cit. on pp. 21, 22).
- Fredrickson A, Megee III R, Tsuchiya H (1970)**. Mathematical models for fermentation processes. *Advances in Applied Microbiology* 13:419–465 (cit. on p. 21).
- Fritzscht FS, Dusny C, Frick O, Schmid A (2012)**. Single-cell analysis in biotechnology, systems biology, and biocatalysis. *Annual Review of Chemical and Biomolecular Engineering* 3:129–155 (cit. on pp. 4, 12, 17, 47).
- Fuller WA (2009)**. *Measurement error models*. Vol. 305. John Wiley & Sons (cit. on p. 141).
- Gartland K, Bruschi F, Dundar M, Gahan P, Viola Magni M, Akbarova Y (2013)**. Progress towards the 'Golden Age' of biotechnology. *Current Opinion in Biotechnology* 24:S6–S13 (cit. on p. 1).
- Gelves R, Dietrich A, Takors R (2014)**. Modeling of gas–liquid mass transfer in a stirred tank bioreactor agitated by a Rushton turbine or a new pitched blade impeller. *Bioprocess and biosystems engineering* 37.3:365–375 (cit. on p. 14).
- Gernaey KV, Lantz AE, Tufvesson P, Woodley JM, Sin G (2010)**. Application of mechanistic models to fermentation and biocatalysis for next-generation processes. *Trends in Biotechnology* 28.7:346–354 (cit. on p. 21).
- Glazko GV, Emmert-Streib F (2009)**. Unite and conquer: univariate and multivariate approaches for finding differentially expressed gene sets. *Bioinformatics* 25.18:2348–2354 (cit. on p. 30).
- Goeman JJ, Van De Geer SA, De Kort F, Van Houwelingen HC (2004)**. A global test for groups of genes: testing association with a clinical outcome. *Bioinformatics* 20.1:93–99 (cit. on pp. 29, 31, 42, 44, 105, 107, 108).
- Gombert AK, Nielsen J (2000)**. Mathematical modelling of metabolism. *Current Opinion in Biotechnology* 11.2:180–186 (cit. on p. 21).
- Gopal GJ, Kumar A (2013)**. Strategies for the production of recombinant protein in *Escherichia coli*. *The Protein Journal* 32.6:419–425 (cit. on pp. 5, 133).
- Green MR, Sambrook J (2012)**. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press New York (cit. on pp. 26, 135, 137).
- Gucker FT, O'Konski C (1949)**. Electronic methods of counting aerosol particles. *Chemical Reviews* 44.2:373–388 (cit. on p. 18).
- Hanahan D, Meselson M (1980)**. Plasmid screening at high colony density. *Gene* 10.1:63–67 (cit. on p. 136).

- Harshey RM (2003)**. Bacterial motility on a surface: many ways to a common goal. *Annual Reviews in Microbiology* 57.1:249–273 (cit. on pp. 43, 110).
- Hatzis C, Porro D (2006)**. Morphologically-structured models of growing budding yeast populations. *Journal of Biotechnology* 124.2:420–438 (cit. on p. 22).
- Heipieper HJ, Neumann G, Cornelissen S, Meinhardt F (2007)**. Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems. *Applied Microbiology and Biotechnology* 74.5:961–973 (cit. on p. 124).
- Helmstetter C (1996)**. Timing of synthetic activities in the cell cycle. *Escherichia coli and Salmonella: cellular and molecular biology*. Ed. by Neidhardt FC. Vol. 2. ASM press Washington, DC (cit. on pp. 55–57, 71, 115, 121, 122, 125).
- Helmstetter CE, Pierucci O (1976)**. DNA synthesis during the division cycle of three substrains of *Escherichia coli* B/r. *Journal of Molecular Biology* 102.3:477–486 (cit. on pp. 55, 125).
- Henson MA (2003)**. Dynamic modeling of microbial cell populations. *Current Opinion in Biotechnology* 14.5:460–467 (cit. on pp. 22, 23).
- Herbert D, Elsworth R, Telling R (1956)**. The continuous culture of bacteria; a theoretical and experimental study. *Journal of General Microbiology* 14.3:601–622 (cit. on pp. 16, 127).
- Herrmann H, Janke D, Krejsa S, Kunze I (1987)**. Involvement of the plasmid pPGH1 in the phenol degradation of *Pseudomonas putida* strain H. *FEMS Microbiology Letters* 43.2:133–137 (cit. on p. 8).
- Hewitt CJ, Nebe-von Caron G, Nienow AW, McFarlane CM (1999)**. The use of multi-parameter flow cytometry to compare the physiological response of *Escherichia coli* W3110 to glucose limitation during batch, fed-batch and continuous culture cultivations. *Journal of Biotechnology* 75.2:251–264 (cit. on pp. 3, 45, 110).
- Hill NS, Kadoya R, Chattoraj DK, Levin PA (2012)**. Cell size and the initiation of DNA replication in bacteria. *PLoS Genetics* 8.3:e1002549 (cit. on p. 59).
- Hjort K, Bernander R (2001)**. Cell cycle regulation in the hyperthermophilic crenarchaeon *Sulfolobus acidocaldarius*. *Molecular Microbiology* 40.1:225–234 (cit. on pp. 57, 126).
- Hoffmann F, Rinas U (2004)**. Stress induced by recombinant protein production in *Escherichia coli*. *Physiological Stress Responses in Bioprocesses*. Springer (cit. on p. 132).
- Hōrak R, Kivisaar M (1998)**. Expression of the transposase gene *tnpA* of Tn4652 is positively affected by integration host factor. *Journal of Bacteriology* 180.11:2822–2829 (cit. on p. 148).
- Hoskisson PA, Hobbs G (2005)**. Continuous culture-making a comeback? *Microbiology* 151.10:3153–3159 (cit. on pp. 15, 16).
- Huang KH, Durand-Heredia J, Janakiraman A (2013)**. FtsZ ring stability: of bundles, tubules, crosslinks, and curves. *Journal of Bacteriology* 195.9:1859–1868 (cit. on p. 125).
- Jahn M, Seifert J, von Bergen M, Schmid A, Bühler B, Müller S (2012)**. Subpopulation-proteomics in prokaryotic populations. *Current Opinion in Biotechnology* (cit. on pp. 12, 13, 17, 104, 105).

- Jahn M, Seifert J, Hübschmann T, von Bergen M, Harms H, Müller S (2013)**. Comparison of preservation methods for bacterial cells in cytomics and proteomics. *Journal of Integrated OMICS* 3.1:25–33 (cit. on pp. 28, 29, 41, 103, 104, 117).
- Jahn M, Vorpahl C, Türkowsky D, Lindmeyer M, Bühler B, Harms H, Müller S (2014)**. Accurate determination of plasmid copy number of flow-sorted cells using droplet digital PCR. *Analytical Chemistry* (cit. on p. 39).
- Jana S, Deb J (2005)**. Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Applied Microbiology and Biotechnology* 67.3:289–298 (cit. on p. 133).
- Jandt U, Barradas OP, Pörtner R, Zeng AP (2014)**. Mammalian cell culture synchronization under physiological conditions and population dynamic simulation. *Applied Microbiology and Biotechnology* 98.10:4311–4319 (cit. on p. 69).
- Jehmlich N, Hübschmann T, Gesell Salazar M, Völker U, Benndorf D, Müller S, von Bergen M, Schmidt F (2010)**. Advanced tool for characterization of microbial cultures by combining cytomics and proteomics. *Applied Microbiology and Biotechnology* 88.2:575–584 (cit. on pp. 19, 104, 117).
- Jiménez JI, Miñambres B, Garcia JL, Díaz E (2002)**. Genomic analysis of the aromatic catabolic pathways from *Pseudomonas putida* KT2440. *Environmental Microbiology* 4.12:824–841 (cit. on p. 8).
- Johnson A, O’Donnell M (2005)**. Cellular DNA replicases: components and dynamics at the replication fork. *Annual Review of Biochemistry* 74:283–315 (cit. on pp. 59, 123).
- Kamentsky LA, Melamed MR, Derman H (1965)**. Spectrophotometer: new instrument for ultrarapid cell analysis. *Science* 150.3696:630–631 (cit. on p. 18).
- Kanehisa M, Goto S (2000)**. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 28.1:27–30 (cit. on p. 30).
- Kazmierczak BI, Hendrixson DR (2013)**. Spatial and numerical regulation of flagellar biosynthesis in polarly flagellated bacteria. *Molecular Microbiology* 88.4:655–663 (cit. on p. 152).
- Keasling J, Kuo H, Vahanian G (1995)**. A Monte Carlo simulation of the *Escherichia coli* cell cycle. *Journal of Theoretical Biology* 176.3:411–430 (cit. on pp. 55, 56, 121, 122).
- Khatri P, Draghici S, Ostermeier GC, Krawetz SA (2002)**. Profiling gene expression using onto-express. *Genomics* 79.2:266–270 (cit. on p. 31).
- Khatri P, Sirota M, Butte AJ (2012)**. Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Computational Biology* 8.2:e1002375 (cit. on pp. 30, 31).
- Kim J, Park W (2014)**. Oxidative stress response in *Pseudomonas putida*. *Applied Microbiology and Biotechnology* 98.16:6933–6946 (cit. on pp. 61, 133).
- Kim S, Dallmann HG, McHenry CS, Mariani KJ (1996)**. Coupling of a replicative polymerase and helicase: a τ -DnaB interaction mediates rapid replication fork movement. *Cell* 84.4:643–650 (cit. on pp. 59, 123).
- Kitano H (2002)**. Systems biology: A brief overview. *Science* 295.5560:1662–1664 (cit. on p. 17).

- Kogoma T (1997)**. Stable DNA replication: interplay between DNA replication, homologous recombination, and transcription. *Microbiology and Molecular Biology Reviews* 61.2:212–238 (cit. on p. 124).
- Kooijman S, Muller E, Stouthamer A (1992)**. Microbial growth dynamics on the basis of individual budgets. *Quantitative Aspects of Growth and Metabolism of Microorganisms*. Springer (cit. on pp. 64, 143).
- Kubitschek H, Newman C (1978)**. Chromosome replication during the division cycle in slowly growing, steady-state cultures of three *Escherichia coli* B/r strains. *Journal of Bacteriology* 136.1:179–190 (cit. on pp. 55, 125).
- Langmead B, Trapnell C, Pop M, Salzberg SL, et al. (2009)**. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* 10.3:R25 (cit. on p. 118).
- Lapin A, Müller D, Reuss M (2004)**. Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler-Lagrange methods: Traveling along the lifelines of single cells. *Industrial and Engineering Chemistry Research* 43.16:4647–4656 (cit. on p. 23).
- Lara AR, Galindo E, Ramirez OT, Palomares LA (2006)**. Living with heterogeneities in bioreactors. *Molecular Biotechnology* 34.3:355–381 (cit. on pp. 4, 14, 47).
- Laub MT, McAdams HH, Feldblyum T, Fraser CM, Shapiro L (2000)**. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* 290.5499:2144–2148 (cit. on p. 42).
- Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, et al. (2008)**. MIFlowCyt: the minimum information about a flow cytometry experiment. *Cytometry Part A* 73.10:926–930 (cit. on p. 117).
- Lee SY, Mattanovich D, Villaverde A (2012)**. Systems metabolic engineering, industrial biotechnology and microbial cell factories. *Microbial Cell Factories* 11:156 (cit. on pp. 1, 5).
- Lencastre Fernandes R, Nierychlo M, Lundin L, Pedersen AE, Puentes Tellez P, Dutta A, Carlquist M, Bolic A, Schäpper D, Brunetti AC, Helmark S, Heins AL, Jensen A, Nopens I, Rottwitt K, Szita N, van Elsas J, Nielsen P, Martinussen J, Sørensen S, Lantz AE, Germaey KV (2011)**. Experimental methods and modeling techniques for description of cell population heterogeneity. *Biotechnology Advances* 29.6:575–599 (cit. on pp. 3, 4, 11, 14, 18, 22–24, 37, 47, 101).
- Lieder S, Jahn M, Seifert J, von Bergen M, Müller S, Takors R (2014)**. Subpopulation-proteomics reveal growth rate, but not cell cycling, as a major impact on protein composition in *Pseudomonas putida* KT2440. *Applied Microbiology and Biotechnology Express* 4:71 (cit. on p. 44).
- Lindås AC, Bernander R (2013)**. The cell cycle of archaea. *Nature Reviews Microbiology* 11.9:627–638 (cit. on pp. 57, 126).

- Lindmo T (1982)**. Kinetics of protein and DNA synthesis studied by mathematical modelling of flow cytometric protein and DNA histograms. *Cell Proliferation* 15.2:197–211 (cit. on pp. 52, 111, 128).
- Luo W, Friedman M, Shedden K, Hankenson K, Woolf P (2009)**. GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* 10.1:161 (cit. on pp. 29, 31, 42, 44, 104, 107, 108).
- Maaløe O, Kjeldgaard NO (1966)**. Control of macromolecular synthesis; a study of DNA, RNA, and protein synthesis in bacteria. *W.A. Benjamin, New York* (cit. on pp. 45, 110).
- Mantzaris NV, Daoutidis P, Srienc F (2001a)**. Numerical solution of multi-variable cell population balance models: I. Finite difference methods. *Computers & Chemical Engineering* 25.11:1411–1440 (cit. on p. 23).
- **(2001b)**. Numerical solution of multi-variable cell population balance models. II. Spectral methods. *Computers & Chemical Engineering* 25.11:1441–1462 (cit. on p. 23).
- **(2001c)**. Numerical solution of multi-variable cell population balance models. III. Finite element methods. *Computers & Chemical Engineering* 25.11:1463–1481 (cit. on p. 23).
- Mantzaris NV, Srienc F, Daoutidis P (2002)**. Nonlinear productivity control using a multi-staged cell population balance model. *Chemical Engineering Science* 57.1:1–14 (cit. on p. 22).
- Martínez-García E, de Lorenzo V (2011a)**. Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of *Pseudomonas putida* KT2440. *Environmental Microbiology* 13.10:2702–2716 (cit. on pp. 5, 133, 134).
- Martínez-García E, Calles B, Arévalo-Rodríguez M, de Lorenzo V (2011b)**. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiology* 11.1:38 (cit. on p. 133).
- Martínez-García E, Jatsenko T, Kivisaar M, Lorenzo V (2014a)**. Freeing *Pseudomonas putida* KT2440 of its proviral load strengthens endurance to environmental stresses. *Environmental Microbiology* (cit. on pp. 5, 9, 68, 133, 136, 150, 152).
- Martínez-García E, Nickel PI, Chavarría M, Lorenzo V (2014b)**. The metabolic cost of flagellar motion in *Pseudomonas putida* KT2440. *Environmental Microbiology* 16.1:291–303 (cit. on pp. 5, 62, 68, 72, 73, 133, 134, 152).
- Martínez-García E, Nickel PI, Aparicio T, de Lorenzo V (submitted 2014)**. *Pseudomonas* 2.0: Genetic upgrading of *Pseudomonas putida* KT2440 as an enhanced host for heterologous gene expression. *Microbial Cell Factories* (cit. on pp. 62, 72, 136, 141).
- McCarthy DJ, Chen Y, Smyth GK (2012)**. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* 40.10:4288–4297 (cit. on pp. 30, 118).
- Meijnen JP, de Winde JH, Ruijssenaars HJ (2008)**. Engineering *Pseudomonas putida* S12 for efficient utilization of D-xylose and L-arabinose. *Applied and Environmental Microbiology* 74.16:5031–5037 (cit. on p. 101).

- Meijnen JP, Verhoef S, Briedjlal AA, de Winde JH, Ruijssenaars HJ (2011)**. Improved p-hydroxybenzoate production by engineered *Pseudomonas putida* S12 by using a mixed-substrate feeding strategy. *Applied Microbiology and Biotechnology* 90.3:885–893 (cit. on p. 9).
- Michelsen O, de Mattos MJT, Jensen PR, Hansen FG (2003)**. Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r. *Microbiology* 149.4:1001–1010 (cit. on pp. 54, 56, 125).
- Migula W (1894)**. Über ein neues System der Bakterien. *Arbeiten aus dem Bakteriologischen Institut der Technischen Hochschule zu Karlsruhe* 1:235–238 (cit. on p. 7).
- Mitchison JM (1977)**. Cell differentiation in microorganisms, plants and animals. Ed. by Nover L, Mothes K. North-Holland, Amsterdam. Chap. Enzyme synthesis during the cell cycle (cit. on p. 3).
- Mizoguchi H, Mori H, Fujio T (2007)**. *Escherichia coli* minimum genome factory. *Biotechnology and Applied Biochemistry* 46.3:157–167 (cit. on pp. 5, 68, 133, 152).
- Müller S (2007)**. Modes of cytometric bacterial DNA pattern: a tool for pursuing growth. *Cell Proliferation* 40.5:621–639 (cit. on pp. 13, 101, 115, 125, 126).
- Müller S, Babel W (2003)**. Analysis of bacterial DNA pattern as an approach for controlling biotechnological processes. *Journal of Microbiological Methods* 55.3:851–858 (cit. on pp. 13, 18, 48, 57, 100, 106).
- Müller S, Harms H, Bley T (2010)**. Origin and analysis of microbial population heterogeneity in bioprocesses. *Current Opinion in Biotechnology* 21.1:100–113 (cit. on pp. 2–4, 11–14, 18, 19, 37, 39, 100, 101, 115).
- Monod J (1949)**. The growth of bacterial cultures. *Annual Reviews in Microbiology* 3.1:371–394 (cit. on pp. 16, 33, 127).
- Monod J (1950)**. La technique de culture continue: Théorie et applications. (cit. on pp. 15, 127).
- Morigen, Løbner-Olesen A, Skarstad K (2003)**. Titration of the *Escherichia coli* DnaA protein to excess *datA* sites causes destabilization of replication forks, delayed replication initiation and delayed cell division. *Molecular Microbiology* 50.1:349–362 (cit. on pp. 59, 126).
- Müller S, Nebe-von Caron G (2010)**. Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. *FEMS Microbiology Reviews* 34.4:554–587 (cit. on pp. 17, 106).
- Myllykallio H, Lopez P, López-García P, Heilig R, Saurin W, Zivanovic Y, Philippe H, Forterre P (2000)**. Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science* 288.5474:2212–2215 (cit. on p. 125).
- Na D, Kim TY, Lee SY (2010)**. Construction and optimization of synthetic pathways in metabolic engineering. *Current Opinion in Microbiology* 13.3:363–370 (cit. on pp. 68, 152).
- Nahku R, Valgepea K, Lahtvee PJ, Erm S, Abner K, Adamberg K, Vilu R (2010)**. Specific growth rate dependent transcriptome profiling of *Escherichia coli* K12 MG1655 in accelerostat cultures. *Journal of Biotechnology* 145.1:60–65 (cit. on pp. 43, 110).

- Nakazawa T (2002)**. Travels of a *Pseudomonas*, from Japan around the world. *Environmental Microbiology* 4.12:782–786 (cit. on p. 8).
- Nakazawa T, Yokota T (1973)**. Benzoate metabolism in *Pseudomonas putida* (arvilla) mt-2: demonstration of two benzoate pathways. *Journal of Bacteriology* 115.1:262–267 (cit. on pp. 8, 101).
- Nam D, Kim SY (2008)**. Gene-set approach for expression pattern analysis. *Briefings in Bioinformatics* 9.3:189–197 (cit. on p. 31).
- Nanchen A, Schicker A, Sauer U (2006)**. Nonlinear dependency of intracellular fluxes on growth rate in miniaturized continuous cultures of *Escherichia coli*. *Applied and Environmental Microbiology* 72.2:1164–1172 (cit. on pp. 64, 143).
- Nebe-von Caron G, Stephens P, Hewitt C, Powell J, Badley R (2000)**. Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting. *Journal of Microbiological Methods* 42.1:97–114 (cit. on p. 18).
- Nelson KE, Weinel C, Paulsen IT, Dodson RJ, Hilbert H, Martins dos Santos VAP, Fouts DE, Gill SR, Pop M, Holmes M, Brinkac L, Beanan M, DeBoy RT, Daugherty S, Kolonay J, Madupu R, Nelson W, White O, Peterson J, Khouri H, Hance I, Chris Lee P, Holtzapple E, Scanlan D, Tran K, Moazzez A, Utterback T, Rizzo M, Lee K, Kosack D, Moestl D, Wedler H, Lauber J, Stjepandic D, Hoheisel J, Straetz M, Heim S, Kiewitz C, Eisen JA, Timmis KN, Dusterhoft A, Tumbler B, Fraser CM (2002)**. Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. *Environmental Microbiology* 4.12:799–808 (cit. on pp. 8, 56, 101, 121, 133).
- Neumeyer A, Hübschmann T, Müller S, Frunzke J (2013)**. Monitoring of population dynamics of *Corynebacterium glutamicum* by multiparameter flow cytometry. *Microbial Biotechnology* 6.2:157–167 (cit. on pp. 3, 45, 110).
- Nicolaou SA, Gaida SM, Papoutsakis ET (2010)**. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: from biofuels and chemicals, to biocatalysis and bioremediation. *Metabolic Engineering* 12.4:307–331 (cit. on pp. 68, 152).
- Nielsen J, Villadsen J (1992)**. Modelling of microbial kinetics. *Chemical Engineering Science* 47.17:4225–4270 (cit. on p. 21).
- Nikel PI, de Lorenzo V (2012)**. Engineering an anaerobic metabolic regime in *Pseudomonas putida* KT2440 for the anoxic biodegradation of 1, 3-dichloroprop-1-ene. *Metabolic Engineering* (cit. on p. 9).
- Nikel PI, Martínez-García E, de Lorenzo V (2014a)**. Biotechnological domestication of pseudomonads using synthetic biology. *Nature Reviews Microbiology* 12.5:368–379 (cit. on pp. 5, 7, 8, 133).
- Nikel PI, de Lorenzo V (2014b)**. Robustness of *Pseudomonas putida* KT2440 as a host for ethanol biosynthesis. *New Biotechnology* (cit. on pp. 61, 138).

- Nikel PI, Silva-Rocha R, Benedetti I, Lorenzo V (2014c)**. The private life of environmental bacteria: pollutant biodegradation at the single cell level. *Environmental Microbiology* 16.3:628–642 (cit. on p. 2).
- Nogales J, Palsson B, Thiele I (2008)**. A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Systems Biology* 2.1:79 (cit. on pp. 8, 61, 64, 133).
- Novick A, Szilard L (1950)**. Experiments with the chemostat on spontaneous mutations of bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 36.12:708 (cit. on pp. 15, 127).
- Palleroni N (1984)**. Genus *Pseudomonas*. Ed. by Krieg N, Holt J. Vol. 1. Williams & Wilkins (cit. on p. 7).
- Pirt S (1965)**. The maintenance energy of bacteria in growing cultures. *Proceedings of the Royal Society of London. Series B, Biological Sciences* (cit. on pp. 16, 36, 140).
- Poblete-Castro I, Becker J, Dohnt K, dos Santos VM, Wittmann C (2012)**. Industrial biotechnology of *Pseudomonas putida* and related species. *Applied Microbiology and Biotechnology* 93.6:2279–2290 (cit. on pp. 7, 61, 101, 133).
- Pósfai G, Plunkett G, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, De Arruda M, Burland V, Harcum SW, Blattner FR (2006)**. Emergent properties of reduced-genome *Escherichia coli*. *Science* 312.5776:1044–1046 (cit. on pp. 68, 152).
- Preusting H, Hazenberg W, Witholt B (1993)**. Continuous production of poly (3-hydroxyalkanoates) by *Pseudomonas oleovorans* in a high-cell-density, two-liquid-phase chemostat. *Enzyme and Microbial Technology* 15.4:311–316 (cit. on p. 43).
- Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, dos Santos VAM (2008)**. Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Computational Biology* 4.10:e1000210 (cit. on pp. 8, 9, 101).
- Ramkrishna D (2000)**. *Population balances: Theory and applications to particulate systems in engineering*. Academic Press (cit. on p. 22).
- Ramkrishna D, Singh MR (2014)**. Population balance modeling: current status and future prospects. *Annual Reviews of Chemical and Biomolecular Engineering* 5:123–146 (cit. on p. 23).
- Rebnegger C, Graf AB, Valli M, Steiger MG, Gasser B, Maurer M, Mattanovich D (2014)**. In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnology Journal* (cit. on pp. 43, 110).
- Reva ON, Weinel C, Weinel M, Böhm K, Stjepandic D, Hoheisel JD, Tümmeler B (2006)**. Functional genomics of stress response in *Pseudomonas putida* KT2440. *Journal of Bacteriology* 188.11:4079–4092 (cit. on p. 8).

- Robert L, Hoffmann M, Krell N, Aymerich S, Robert J, Doumic M (2014)**. Division in *Escherichia coli* is triggered by a size-sensing rather than a timing mechanism. *BMC Biology* 12.1:17 (cit. on pp. 57, 126).
- Robinson MD, Smyth GK (2007)**. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* 23.21:2881–2887 (cit. on p. 30).
- Robinson MD, Smyth GK (2008)**. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 9.2:321–332 (cit. on p. 30).
- Robinson MD, McCarthy DJ, Smyth GK (2010)**. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26.1:139–140 (cit. on pp. 30, 118).
- Rocha EP (2004)**. The replication-related organization of bacterial genomes. *Microbiology* 150.6:1609–1627 (cit. on p. 125).
- Roels J (1980)**. Application of macroscopic principles to microbial metabolism. *Biotechnology and Bioengineering* 22.12:2457–2514 (cit. on p. 24).
- Rønning ØW, Pettersen EO, Seglen PO (1979)**. Protein synthesis and protein degradation through the cell cycle of human NHIK 3025 cells in vitro. *Experimental Cell Research* 123.1:63–72 (cit. on p. 111).
- Rosano GL, Ceccarelli EA (2014)**. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology* 5 (cit. on p. 151).
- Ruiz JA, de Almeida A, Godoy MS, Mezzina MP, Bidart GN, Méndez BS, Pettinari MJ, Nickel PI (2013)**. *Escherichia coli* redox mutants as microbial cell factories for the synthesis of reduced biochemicals. *Computational and Structural Biotechnology Journal* 3:e201210019 (cit. on p. 133).
- Russell JB (2007)**. The energy spilling reactions of bacteria and other organisms. *Journal of Molecular Microbiology and Biotechnology* 13.1-3:1–11 (cit. on pp. 64, 143).
- Rustici G, Mata J, Kivinen K, Lió P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bähler J (2004)**. Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics* 36.8:809–817 (cit. on p. 42).
- Sauer M, Mattanovich D (2012)**. Construction of microbial cell factories for industrial bioprocesses. *Journal of Chemical Technology and Biotechnology* 87.4:445–450 (cit. on pp. 1, 5, 132).
- Schaechter M, Maaløe O, Kjeldgaard NO (1958)**. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *Journal of General Microbiology* 19.3:592–606 (cit. on pp. 43, 45, 110).
- Schmid A, Dordick J, Hauer B, Kiener A, Wubbolts M, Witholt B (2001)**. Industrial biocatalysis today and tomorrow. *Nature* 409.6817:258–268 (cit. on pp. 8, 9).
- Schmid A, Kortmann H, Dittrich PS, Blank LM (2010)**. Chemical and biological single cell analysis. *Current Opinion in Biotechnology* 21.1:12–20 (cit. on p. 17).

- Schneider D, Lenski RE (2004)**. Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology* 155.5:319–327 (cit. on p. 148).
- Schulze KL, Lipe RS (1964)**. Relationship between substrate concentration, growth rate, and respiration rate of *Escherichia coli* in continuous culture. *Archiv für Mikrobiologie* 48.1:1–20 (cit. on pp. 64, 143).
- Schweder T, Krüger E, Xu B, Jürgen B, Blomsten G, Enfors SO, Hecker M (1999)**. Monitoring of genes that respond to process-related stress in large-scale bioprocesses. *Biotechnology and Bioengineering* 65.2:151–159 (cit. on pp. 4, 12, 14, 47).
- Seto S, Miyata M (1998)**. Cell Reproduction and Morphological Changes in *Mycoplasma capricolum*. *Journal of Bacteriology* 180.2:256–264 (cit. on p. 125).
- Shapiro HM (2000)**. Microbial analysis at the single-cell level: tasks and techniques. *Journal of Microbiological Methods* 42.1:3–16 (cit. on pp. 18, 106).
- Sharma SS, Blattner FR, Harcum SW (2007)**. Recombinant protein production in an *Escherichia coli* reduced genome strain. *Metabolic Engineering* 9.2:133–141 (cit. on pp. 68, 152).
- Sherer E, Tocce E, Hannemann R, Rundell A, Ramkrishna D (2008)**. Identification of age-structured models: Cell cycle phase transitions. *Biotechnology and Bioengineering* 99.4:960–974 (cit. on p. 22).
- Sidoli F, Mantalaris A, Asprey S (2004)**. Modelling of mammalian cells and cell culture processes. *Cytotechnology* 44.1-2:27–46 (cit. on pp. 22, 24).
- Silva-Rocha R, Martínez-García E, Calles B, Chavarría M, Arce-Rodríguez A, de las Heras A, Páez-Espino AD, Durante-Rodríguez G, Kim J, Nickel PI, Platero R, de Lorenzo V (2013)**. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research* 41.D1:D666–D675 (cit. on pp. 5, 133, 136, 145).
- Silva F, Queiroz JA, Domingues FC (2012)**. Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnology Advances* 30.3:691–708 (cit. on pp. 66, 151).
- Singh V (2014)**. Recent advancements in synthetic biology: Current status and challenges. *Gene* 535.1:1–11 (cit. on pp. 5, 61, 132).
- Skarstad K, Steen HB, Boye E (1985)**. *Escherichia coli* DNA distributions measured by flow cytometry and compared with theoretical computer simulations. *Journal of Bacteriology* 163.2:661–668 (cit. on pp. 45, 48, 54–56, 70, 106, 115, 116, 118, 121, 124, 129, 131).
- Skarstad K, Steen HB, Boye E (1983)**. Cell cycle parameters of slowly growing *Escherichia coli* B/r studied by flow cytometry. *Journal of Bacteriology* 154.2:656–662 (cit. on pp. 18, 110).
- Sohn SB, Kim TY, Park JM, Lee SY (2010)**. In silico genome-scale metabolic analysis of *Pseudomonas putida* KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival. *Biotechnology Journal* 5.7:739–750 (cit. on p. 8).

- Soriano E, Borth N, Katinger H, Mattanovich D (1999)**. Flow cytometric analysis of metabolic stress effects due to recombinant plasmids and proteins in *Escherichia coli* production strains. *Metabolic Engineering* 1.3:270–274 (cit. on p. 148).
- Soutourina OA, Bertin PN (2003)**. Regulation cascade of flagellar expression in Gram-negative bacteria. *FEMS Microbiology Reviews* 27.4:505–523 (cit. on pp. 43, 110).
- Spidlen J, Breuer K, Rosenberg C, Kotecha N, Brinkman RR (2012)**. FlowRepository: A resource of annotated flow cytometry datasets associated with peer-reviewed publications. *Cytometry Part A* 81.9:727–731 (cit. on p. 117).
- Srienc F (1999)**. Cytometric data as the basis for rigorous models of cell population dynamics. *Journal of Biotechnology* 71.1-3:233–238 (cit. on pp. 18, 23, 115).
- Stamatakis M (2010)**. Cell population balance, ensemble and continuum modeling frameworks: Conditional equivalence and hybrid approaches. *Chemical Engineering Science* 62.2:1008–1015 (cit. on p. 23).
- Steen HB (2001)**. Flow cytometers for characterization of microorganisms. *Current Protocols in Cytometry*:1–11 (cit. on p. 115).
- Stelling J (2004)**. Mathematical models in microbial systems biology. *Current Opinion in Microbiology* 7.5:513–518 (cit. on p. 21).
- Stokke C, Flåtten I, Skarstad K (2012)**. An easy-to-use simulation program demonstrates variations in bacterial cell cycle parameters depending on medium and temperature. *PLoS One* 7.2:e30981 (cit. on p. 54).
- Studwell PS, O'Donnell M (1990)**. Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme. *Journal of Biological Chemistry* 265.2:1171–1178 (cit. on pp. 58, 123).
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005)**. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102.43:15545–15550 (cit. on p. 30).
- Tatusov RL, Koonin EV, Lipman DJ (1997)**. A genomic perspective on protein families. *Science* 278.5338:631–637 (cit. on pp. 29, 42, 44, 58, 104, 107, 108, 113, 118, 122, 124).
- Teather R, Collins J, Donachie W (1974)**. Quantal behavior of a diffusible factor which initiates septum formation at potential division sites in *Escherichia coli*. *Journal of Bacteriology* 118.2:407–413 (cit. on p. 111).
- Tempest DW (1970)**. The continuous cultivation of microorganisms 1. Theory of the chemostat. *Methods in Microbiology*. Ed. by Norris JR, Ribbons DW. Vol. 2. London: Academic Press Inc. (cit. on pp. 15, 31).

- Theobald U, Mailinger W, Baltes M, Rizzi M, Reuss M (1997)**. In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. *Biotechnology and Bioengineering* 55.2:305–316 (cit. on pp. 102, 103).
- Tian L, Greenberg SA, Kong SW, Altschuler J, Kohane IS, Park PJ (2005)**. Discovering statistically significant pathways in expression profiling studies. *Proceedings of the National Academy of Sciences of the United States of America* 102.38:13544–13549 (cit. on p. 31).
- Timmis KN (2002)**. *Pseudomonas putida*: a cosmopolitan opportunist par excellence. *Environmental Microbiology* 4.12:779–781 (cit. on pp. 7, 8).
- Tracy BP, Gaida SM, Papoutsakis ET (2010)**. Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Current Opinion in Biotechnology* 21.1:85–99 (cit. on p. 18).
- Tyers M, Mann M (2003)**. From genomics to proteomics. *Nature* 422.6928:193–197 (cit. on p. 17).
- Umenhoffer K, Fehér T, Balikó G, Ayaydin F, Pósfai J, Blattner FR, Pósfai G (2010)**. Research reduced evolvability of *Escherichia coli* MDS42, an IS-less cellular chassis for molecular and synthetic biology applications. *Microbial Cell Factories* 9:38 (cit. on pp. 68, 152).
- Unthan S, Grünberger A, van Ooyen J, Gätgens J, Heinrich J, Paczia N, Wiechert W, Kohlheyer D, Noack S (2014)**. Beyond growth rate 0.6: What drives *Corynebacterium glutamicum* to higher growth rates in defined medium. *Biotechnology and Bioengineering* 111.2:359–371 (cit. on pp. 38, 101).
- Vallon T, Glemser M, Malca SH, Scheps D, Schmid J, Siemann-Herzberg M, Hauer B, Takors R (2013)**. Production of 1-octanol from n-octane by *Pseudomonas putida* KT2440. *Chemie Ingenieur Technik* 85.6:841–848 (cit. on p. 143).
- Van Duuren JB, Pucha J, Mars AE, Bücken R, Eggink G, Wittmann C, dos Santos VA (2013)**. Reconciling in vivo and in silico key biological parameters of *Pseudomonas putida* KT2440 during growth on glucose under carbon-limited condition. *BMC Biotechnology* 13.1:93 (cit. on pp. 64, 143).
- Verhoef S, Ballerstedt H, Volkers RJ, de Winde JH, Ruijsenaars HJ (2010)**. Comparative transcriptomics and proteomics of p-hydroxybenzoate producing *Pseudomonas putida* S12: novel responses and implications for strain improvement. *Applied Microbiology and Biotechnology* 87.2:679–690 (cit. on p. 9).
- Villadsen J, Nielsen J, Lidén G (2011)**. *Bioreaction engineering principles*. Springer (cit. on pp. 23, 24, 33, 36).
- Vizcaino-Caston I, Wyre C, Overton TW (2012)**. Fluorescent proteins in microbial biotechnology – new proteins and new applications. *Biotechnology Letters* 34.2:175–186 (cit. on pp. 62, 133).
- Wackett LP (2003)**. *Pseudomonas putida* - a versatile biocatalyst. *Nature Biotechnology* 21.2:136–138 (cit. on p. 8).

- Waegeman H, Soetaert W (2011)**. Increasing recombinant protein production in *Escherichia coli* through metabolic and genetic engineering. *Journal of Industrial Microbiology and Biotechnology* 38.12:1891–1910 (cit. on p. 152).
- Waldbauer JR, Rodrigue S, Coleman ML, Chisholm SW (2012)**. Transcriptome and proteome dynamics of a light-dark synchronized bacterial cell cycle. *PloS one* 7.8:e43432 (cit. on p. 43).
- Wang JD, Levin PA (2009)**. Metabolism, cell growth and the bacterial cell cycle. *Nature Reviews Microbiology* 7.11:822–827 (cit. on p. 115).
- Weart RB, Lee AH, Chien AC, Haeusser DP, Hill NS, Levin PA (2007)**. A metabolic sensor governing cell size in bacteria. *Cell* 130.2:335–347 (cit. on pp. 42, 107).
- Weinel C, Nelson KE, Tümmler B (2002)**. Global features of the *Pseudomonas putida* KT2440 genome sequence. *Environmental Microbiology* 4.12:809–818 (cit. on p. 133).
- West SC (1996)**. The RuvABC proteins and Holliday junction processing in *Escherichia coli*. *Journal of Bacteriology* 178.5:1237 (cit. on p. 124).
- Whitby MC, Vincent SD, Lloyd R (1994)**. Branch migration of Holliday junctions: identification of RecG protein as a junction specific DNA helicase. *The EMBO Journal* 13.21:5220 (cit. on p. 124).
- Wiacek C, Müller S, Benndorf D (2006)**. A cytomic approach reveals population heterogeneity of *Cupriavidus necator* in response to harmful phenol concentrations. *Proteomics* 6.22:5983–5994 (cit. on pp. 48, 116, 125).
- Wierckx NJ, Ballerstedt H, de Bont JA, Wery J (2005)**. Engineering of solvent-tolerant *Pseudomonas putida* S12 for bioproduction of phenol from glucose. *Applied and Environmental Microbiology* 71.12:8221–8227 (cit. on p. 9).
- Winder CL, Lanthaler K (2011)**. The use of continuous culture in systems biology investigations. *Methods in Enzymology* 500:261–275 (cit. on p. 127).
- Wittenberg C, Reed SI (2005)**. Cell cycle-dependent transcription in yeast: promoters, transcription factors, and transcriptomes. *Oncogene* 24.17:2746–2755 (cit. on p. 42).
- Wong MS, Wu S, Causey TB, Bennett GN, San KY (2008)**. Reduction of acetate accumulation in *Escherichia coli* cultures for increased recombinant protein production. *Metabolic Engineering* 10.2:97–108 (cit. on pp. 66, 151).
- Wu J (2004)**. Global analysis of nutrient control of gene expression in *Saccharomyces cerevisiae* during growth and starvation. *Proceedings of the National Academy of Sciences of the United States of America* 4.101:3148–3153 (cit. on p. 17).
- Yuste L, Hervás AB, Canosa I, Tobes R, Jiménez JI, Nogales J, Pérez-Pérez MM, Santero E, Díaz E, Ramos JL, de Lorenzo V, Rojo F (2006)**. Growth phase-dependent expression of the *Pseudomonas putida* KT2440 transcriptional machinery analysed with a genome-wide DNA microarray. *Environmental Microbiology* 8.1:165–177 (cit. on p. 101).

- Zenobi R (2013)**. Single-cell metabolomics: analytical and biological perspectives. *Science* 342.6163:1243259 (cit. on p. 45).
- Zhen D, Liu H, Wang SJ, Zhang JJ, Zhao F, Zhou NY (2006)**. Plasmid-mediated degradation of 4-chloronitrobenzene by newly isolated *Pseudomonas putida* strain ZWL73. *Applied Microbiology and Biotechnology* 72.4:797–803 (cit. on p. 8).

A. Manuscript I

Population heterogeneity occurring in industrial microbial bioprocesses is regarded as a putative effector causing performance loss in large scale. While the existence of subpopulations is a commonly accepted fact, their appearance and impact on process performance still remains rather unclear. During cell cycling, distinct subpopulations differing in cell division state and DNA content appear which contribute individually to the efficiency of the bioprocess. To identify stressed or impaired subpopulations, we analyzed the interplay of growth rate, cell cycle and phenotypic profile of subpopulations by using flow cytometry and cell sorting in conjunction with mass spectrometry based global proteomics. Adjusting distinct growth rates in chemostats with the model strain *P. putida* KT2440, cells were differentiated by DNA content reflecting different cell cycle stages. The proteome of separated subpopulations at given growth rates was found to be highly similar, while different growth rates caused major changes of the protein inventory with respect to e.g. carbon storage, motility, lipid metabolism and the translational machinery. In conclusion, cells in various cell cycle stages at the same growth rate were found to have similar to identical proteome profiles showing no significant population heterogeneity on the proteome level. In contrast, the growth rate clearly determines the protein composition and therefore the metabolic strategy of the cells.

This chapter has been published as: **Sarah Lieder, Michael Jahn, Jana Seifert, Martin von Bergen, Susann Müller, Ralf Takors** (2014) Subpopulation-proteomics reveal growth rate, but not cell cycling, as a major impact on protein composition in *Pseudomonas putida* KT2440. *Applied Microbiology and Biotechnology Express* 4:71

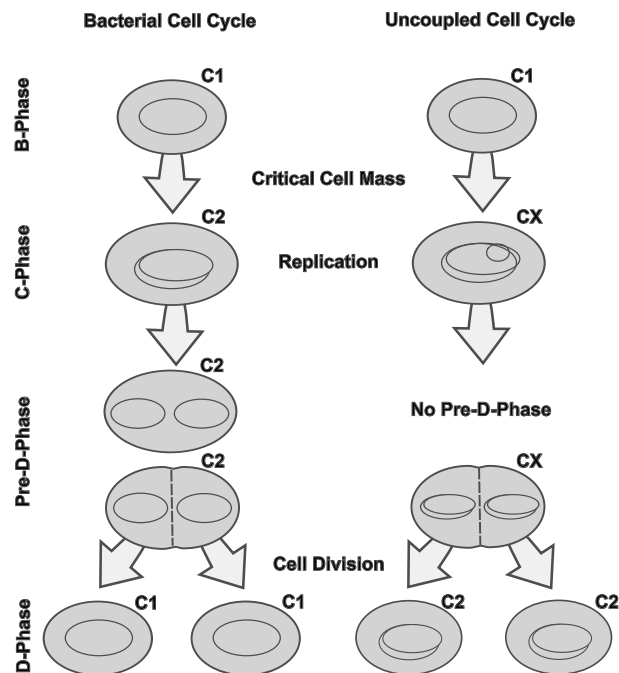


Figure A.1.: Schematic overview of the bacterial cell cycle. The bacterial cell cycle can be divided into B, C, pre-D and D phases constituting a defined order within one generation time. Under unlimited growth conditions, some bacterial species are capable of accelerating proliferation by uncoupling DNA synthesis from division. As a result, a new round of DNA replication is initiated before the completion of the previous round (Cooper, 1991; Müller et al., 2010).

A.1. Introduction

Commonly applied assumptions consider microbial populations in bioreactors as uniform, thus leveling individual properties of subpopulations to averages. However, it is increasingly accepted that clonal microbial cultures comprise individuals that are not identical, differing in terms of DNA content and cell physiology (Brehm-Stecher et al., 2004; Delvigne et al., 2014). Heterogeneity of clonal microbial cultures may result from several distinct sources, either from internal biological origins, such as mutations, cell cycle decisions and age distribution, or from ‘external’ technical factors (Avery, 2006; Müller et al., 2010). Notably, external factors interact with biological properties, yielding the superimposition of both impacts in the population. Here, we shed light on the impact of two key players in the origin of population heterogeneity, the growth rate and the cell cycle.

Traditionally, the cell cycle is suggested to play a role in the development of population heterogeneity within clonal populations (Müller et al., 2010). A short summary of the sequence of cell cycle phases can be found in Figure A.1. The bacterial cell cycle was described for *Escherichia coli* comprising the B-Phase, which is defined as the time between division and start of replication, the replication phase (C-Phase), the pre-D-Phase (an interphase between the C- and D-Phase) and the division phase (D-Phase) (Cooper, 1991; Müller et al., 2003). Furthermore, under optimal growth

conditions accelerated proliferation (also called ‘multifork DNA-replication’) can be monitored: new rounds of DNA replication may be initiated before a previous round is completed, putatively providing another source of heterogeneity (Bley, 1990; Müller, 2007).

It is suspected, that product-biosynthesis of biotechnological interesting compounds occurs in dependency of the cell cycle, e.g. only within the stochastic B- and pre-D-phases, when cells are neither replicating nor dividing (Müller et al., 2010). Ackermann et al. (1995) described for *Methylobacterium rhodesianum* that products like polyhydroxyalkanoates (PHAs) accumulate only when cells comprise a certain chromosome number. This phenomenon was found to occur at off-cell-cycling stages. In microbial biotechnology, heterogeneity caused by cell cycling may cause inefficiently producing subpopulations and could have significant impact on the overall process performance (Lencastre Fernandes et al., 2011). Here, we aim to investigate if the protein inventory of a cell, which is related to its metabolic activity, is dependent on cell cycle stages and how growth rates may influence both, protein composition and cell cycling.

P. putida KT2440 was used as a model organism owing to its numerous qualities as an expression host, such as safety (Bagdasarian et al., 1981; Nakazawa et al., 1973), fast growth, a fully sequenced genome (Nelson et al., 2002) and high stress tolerance (dos Santos et al., 2004). Together with simple nutrient demand, the potential to regenerate redox cofactors at a high rate (Blank et al., 2008) and its amenability to genetic manipulation, *P. putida* is an ideal host for heterologous gene expression (Meijnen et al., 2008). With the advance of genome-wide pathway modeling (Puchałka et al., 2008) and ‘omics techniques, the way for systems-wide engineering strategies was paved to turn *P. putida* into a flexible cell factory chassis (Yuste et al., 2006). Consequently, *P. putida* is more and more explored and already successfully used for numerous industrial applications (Poblete-Castro et al., 2012; Puchałka et al., 2008).

In our study, we applied continuous cultivations under controlled growth conditions at defined growth rates. While (fed-) batch approaches are characterized by steadily changing environmental conditions such as media composition, steady-state modes of a chemostat, where cells are cultivated with a pre-installed growth rate, are defined by environmental conditions that remain unchanged (Carlquist et al., 2012). Notably, (fed-) batch cultures usually represent a mixture of cells growing with different speed as a consequence of changing cultivating conditions (Unthan et al., 2014). Investigating a wide spectrum of growth rates with chemostat cultivation and sampling at steady state conditions gave a specific and unmasked view on the influence of the growth rate on population characteristics. Features like DNA content of the cells, protein composition and adenylate energy charge measurements were included in the study. Additionally, subpopulations with different DNA content were sorted at growth rates 0.1 h^{-1} , 0.2 h^{-1} and 0.7 h^{-1} and analyzed for their proteome composition. Summarizing, we investigated if cell cycling subpopulations at the same growth rate were independent and different from each other on the level of the metabolic pathways, e.g. whether slow growing cells with longer cell cycling phases might specialize between proliferation

and production phases. In addition, we wanted to clarify if cells invest into different protein species under rising growth rates.

A.2. Materials and Methods

Bacterial strains and cultivation conditions

Chemicals were purchased from Fluka, St. Gallen, Switzerland. Experiments were performed with *P. putida* KT2440 (ATCC 47054) cells originating from a single colony stored in a working cell bank at -70°C . Cells were cultivated in M12 minimal salt medium containing 2.2 g L^{-1} $(\text{NH}_4)_2\text{SO}_4$, 0.4 g L^{-1} $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.04 g L^{-1} $\text{CaCl}_2 \cdot \text{H}_2\text{O}$, 0.02 g L^{-1} NaCl , 2 g L^{-1} KH_2PO_4 and trace elements (2 mg L^{-1} $\text{ZnSO}_4 \cdot \text{H}_2\text{O}$, 1 mg L^{-1} $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 15 mg L^{-1} $\text{Na}_3\text{-citrate} \cdot 2\text{H}_2\text{O}$, 1 mg L^{-1} $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.02 mg L^{-1} $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$, 0.03 mg L^{-1} $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$, 0.3 mg L^{-1} H_3BO_3 , 10 mg L^{-1} $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$).

A shake flask preculture (150 mL) was started from a minimal medium working cell bank (8.5 mL) with a glucose concentration of 5 g L^{-1} . At mid-exponential growth phase, the preculture was used to inoculate the bioreactor (KLF 3.7 L, Ser. No. 10819, Bioengineering AG, Wald, Switzerland) to reach a final working volume of 1.5 L. Before inoculation, the cultivation conditions were set to 30°C , a stirrer speed of 700 rpm, a pressure of 0.5 bar and an aeration of 2 L min^{-1} sterile filtered ambient air. The pH was set and maintained at pH 7 with 25% (v/v) NH_4OH . Exhaust gas composition (Blue Sense CO_2 and O_2 , (DCP- CO_2 DCP- O_2 , Blue Sense gas sensor GmbH, Herten, Germany), dissolved oxygen and pH in the liquid phase (Ingold, Mettler Toledo GmbH, Giessen, Germany) were monitored online. After glucose depletion, the batch cultivation was continued as a chemostat. At steady state conditions, the dilution rate equals the specific growth rate μ in a chemostat set up. Each steady state dilution rate (and therefore growth rate) and environmental condition was kept for 5 residence times. The dilution rate was adjusted by feeding at a defined flow rate. Weight gain of the reactor was monitored and a harvest pump was started at a weight gain of 10 g. Additionally, the dilution rate was checked manually by measuring the mass of the harvest outflow within a timespan of one hour before sampling. Steady state was evaluated online via exhaust air analysis. Chemostat cultivations were performed in three individual biological replicates.

Determination of the adenylate energy charge

The adenylate energy charge (AEC) value mirrors the cellular energy status (Atkinson et al., 1967) and can be assessed as follows: Biocatalytic reactions inside the cells were stopped with 35% (w/v) HClO_4 . 4 mL biosuspension was taken directly into 1 mL of precooled (-20°C) HClO_4 solution on ice and mixed immediately (Theobald et al., 1997). The sample was shaken at 4°C for 15 min in an

overhead rotation shaker. Afterwards, the solution was neutralized on ice by fast addition of 1 mL 1 M K_2HPO_4 and 0.9 mL 5 M KOH (Buchholz et al., 2001). The neutral solution was centrifuged at 4 °C and 4,000 x *g* for 10 min to remove cell debris, precipitated protein and potassium perchlorate. The supernatant was kept at −20 °C for batch high pressure liquid chromatography (HPLC) measurements. At each sampling time, the biosuspension sample and a filtrated sample without cells was treated according to the above described procedure. Nucleotide analysis was performed by reversed phase ion pair HPLC (Theobald et al., 1997). The HPLC system (Agilent Technologies, Waldbronn, Germany) consisted of an Agilent 1200 series autosampler, an Agilent 1200 series Binary Pump SL, an Agilent 1200 series thermostated column compartment, and an Agilent 1200 series diode array detector set at 260 and 340 nm. The nucleotides were separated and quantified on an RP-C-18 column that was combined with a guard column (Supelcosil LC-18-T; 15 cm x 4.6 mm, 3 μm packing and Supelguard LC-18-T replacement cartridges, 2 cm; Supelco, Bellefonte, USA) at a flow rate of 1 ml/min. A gradient elution method (Cserjan-Puschmann et al., 1999) was adapted and performed with two mobile phases, buffer A (0.1 M KH_2PO_4/K_2HPO_4 , with 4 mM tetrabutylammonium sulfate and 0.5 % (v/v) methanol, pH 6.0) and (ii) solvent B (70 % (v/v) buffer A and 30 % (v/v) methanol, pH 7.2). The following gradient programs were implemented: 100 % (v/v) buffer A from 0 min to 3.5 min, increased to 100 % (v/v) B until 43.5 min, remaining at 100 % (v/v) B until 51 min, decreased to 100 % (v/v) A until 56 min and remaining at 100 % (v/v) A until 66 min.

The AEC is calculated according to Atkinson et al. (1967):

$$AEC = ([ATP] + 0.5 \cdot [ADP]) / ([AMP] + [ADP] + [ATP])$$

Sample preparation and staining for flow Cytometry

Samples for flow cytometry were washed with PBS, resuspended in cryo-protective solution (15 % (v/v) Glycerol in PBS according to Jahn et al. (2013)) and stored at −20 °C. Deep-frozen cell samples were thawed on ice and centrifuged for 2 min min at 8,000 x *g* and 4 °C to remove the cryo-protective solution. The supernatant was discarded, the cells were resuspended in ice cold PBS and adjusted to an optical density of $OD_{600nm} = 0.05$ in 2 mL volume. For DNA staining, the cells were centrifuged, taken up in 1 mL permeabilization buffer (0.1 M citric acid, 5 g L^{−1} Tween 20), incubated for 10 min on ice, centrifuged again and the supernatant was removed. Finally, cells were resuspended in 2 mL ice cold staining buffer (0.68 μM DAPI, 0.1 M Na_2HPO_4), filtered through a Partec CellTrics mesh (Partec, Germany) with 30 μm pore size and stored on ice until analysis.

Flow cytometry and cell sorting

Flow cytometry was performed on biological duplicates. For each biological replicate two technical replicates were investigated using a MoFlo cell sorter (Beckman-Coulter, USA) as described before (Jahn et al., 2012; Jehmlich et al., 2010). Forward scatter (FSC) and side scatter signals (SSC) were acquired using blue laser excitation (488 nm, 400 mW) and a bandpass filter of 488/10 nm together with a neutral density filter of 2.0 for emission. The DAPI fluorescence was recorded using a multi-line UV laser for excitation (333-365 nm, 100 mW) and a bandpass filter of 450 ± 30 nm for emission. Cells were sorted at the most accurate mode (single cell, one drop) with a sorting speed of $4,000 \text{ s}^{-1}$ and a sample chamber cooled to 4°C . For cell sorting a total number of 5×10^6 cells per replicate was directly sorted on a filter well plate (LoProdyneTM membrane with $0.45 \mu\text{m}$ pore size, Nunc, Germany) and the residual buffer was constantly drawn off by an exhaust pump. After sorting, the filter membrane was washed three times with $200 \mu\text{L}$ PBS, air dried and stored at -20°C for further analysis.

Identification of proteins by LC-MS-MS

For quantitative proteomics, the filter membrane was cut into smaller pieces and treated by trypsin for whole cell proteolytic digestion as described in Jahn et al. (2013). The obtained peptide solution was purified using the ZipTip protocol (Millipore, USA), dried in a vacuum concentrator at 30°C and finally taken up in $20 \mu\text{L}$ 0.1% (w/v) formic acid. The solution was separated by nano-ultra performance liquid chromatography and measured by an LTQ Orbitrap XL (Thermo Fisher Scientific, Germany) as described in Jahn et al. (2013).

Data analysis

Mass spectra were analyzed by MaxQuant v1.2.2.5 (Cox et al., 2008) for protein identification and label-free quantification with the genome database of *P. putida* KT2440 and the settings given in Jahn et al. (2013). The label-free quantification (LFQ) values were used for further data analysis and can be found in the supplementary dataset 1 (section A.6 ‘Supplemental material’). The mean, standard deviation and relative quantity of replicates in relation to the reference population (RP, $\mu = 0.2 \text{ h}^{-1}$, mean of two biological replicates) was calculated. The RP was sorted in order to exclude influences of the sorting procedure on the proteomic content. Unsorted cells of the 0.2 h^{-1} grown population were used as an unaffected control population (CP). Student’s *t*-test was performed for significance testing ($p < 0.05$) of single proteins. Proteins were annotated using COG (clusters of orthologous groups) (Tatusov et al., 1997) and clustered in two hierarchical levels of metabolic pathways (‘metabolism’, ‘pathway’). Protein clusters were tested for significant changes using the R Bioconductor (www.bioconductor.org) packages GAGE (Luo et al., 2009)

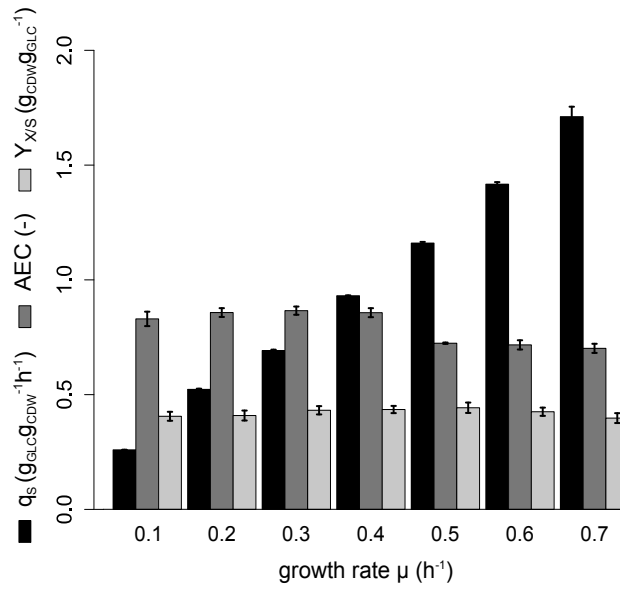


Figure A.2.: Summary of the physiological state of the average population. The specific glucose uptake rate (q_s , $\text{g}_{\text{GLC}} \text{g}_{\text{CDW}}^{-1} \text{h}^{-1}$, black bars), the adenylate energy charge (AEC, dark grey bars) and the biomass yield ($Y_{X/S}$, $\text{g}_{\text{CDW}} \text{g}_{\text{GLC}}^{-1}$, light grey bars) were measured at steady state conditions for different growth rates μ (h^{-1}). The growth rate was stepwise increased until a wash-out of the cells was monitored. Concentrations of cell dry weight (CDW), glucose (GLC) and the AEC were measured offline, sampling after 5 residence times of one specific growth rate ($0.1 \leq \mu (\text{h}^{-1}) \leq 0.7$). Error bars show the standard deviation between three biological replicate cultivations.

and GlobalTest (Goeman et al., 2004), setting $p < 0.05$ and a relative fold change (FC) of 1.5 ($\log_2 \text{FC} = 0.58$) as thresholds. Hierarchical groups were visualized using a color-coded circular treemap (Jahn et al., 2012).

A.3. Results

Subpopulation dynamics of *P. putida* KT2440 were analyzed in a wide range from slow growth rates starting at $\mu = 0.1 \text{ h}^{-1}$ to high growth rates of up to $\mu = 0.7 \text{ h}^{-1}$. At growth rates higher than $\mu = 0.7 \text{ h}^{-1}$, wash out of the culture was observed, meaning that the maximal growth rate was exceeded and cells could not reproduce fast enough to keep the population density constant. For this reason, $\mu = 0.7 \text{ h}^{-1}$ was the highest growth rate investigated in this study. The physiological and the energetic state of the averaged cell population was analyzed by biomass/substrate yield ($Y_{X/S}$), biomass specific substrate uptake rates (q_s), and adenylate energy charge measurements (AEC), each measured at steady state growth conditions (Figure A.2). Observed stable carbon dioxide emission rates served as the criterion to qualify the achievement of steady-state cultivation conditions.

The yield of biomass on glucose increased gradually by 10% from $\mu = 0.1 \text{ h}^{-1}$ to $\mu = 0.5 \text{ h}^{-1}$. Further rise of the growth rate resulted in yield reductions, returning to the level at $\mu = 0.1 \text{ h}^{-1}$

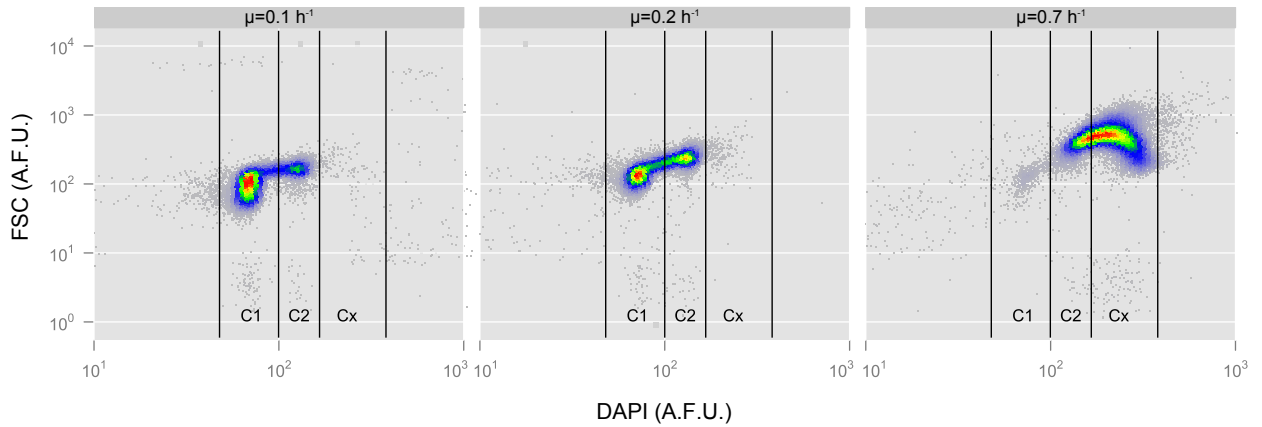


Figure A.3.: Dot plots of DNA content (DAPI, in arbitrary fluorescence units (A.F.U.)) versus forward scatter (FSC, in A.F.U.) at different growth rates 0.1 h^{-1} , 0.2 h^{-1} and 0.7 h^{-1} . The dataset of the biological replicate can be found in section A.6 ‘Supplementary material’ (Figure A.6). The DNA content and the forward scatter increased with increasing growth rate. The indicated gates (C1, C2, Cx) were used for sorting 5×10^6 cells per subpopulation for further mass spectrometric analysis.

(-10%). The energetic capacity of the cells can be estimated via AEC, taking the relative contribution of all three phosphorylated forms of adenine into account. The AEC was found to be stable with increasing growth rate until $\mu = 0.4 \text{ h}^{-1}$. Further increasing the growth rate resulted in a reduction of the AEC level by -18% ($p\text{-value} < 0.01$) which was almost the same at maximum growth. The specific glucose uptake rate q_s was increasing linearly with increasing growth rate.

To be able to distinguish between subpopulations, flow cytometry was proven to be a suitable tool shedding light on the dynamics of single cells within a heterogeneous microbial population (Cooper, 1991; Müller et al., 2003; Shapiro, 2000; Skarstad et al., 1985). Here, the DNA content was monitored via flow cytometry in addition to forward scattering (FSC) giving relative information about cell size (Müller et al., 2010) (Figure A.3). The dataset of the biological replicate can be found in section A.6 ‘Supplementary material’ (Figure A.6). The subpopulation analysis revealed that the major differential parameter was the alteration of DNA content as distinguished by flow cytometry. Three subpopulations could be identified in total: cells containing a single chromosome equivalent (C1), two chromosome equivalents (C2) and cells with more than two chromosome equivalents (Cx) (Figure A.3). Population composition with respect to DNA content varied clearly as a function of growth rates. At $\mu = 0.1 \text{ h}^{-1}$, $82.0 \pm 0.3 \%$ of cells contained a single chromosome equivalent, while only $18.0 \pm 0.2 \%$ contained a double chromosome equivalent content. No Cx subpopulation could be detected. On the contrary, at the high growth rate of $\mu = 0.7 \text{ h}^{-1}$ only $1.4 \pm 0.8 \%$ of cells belonged to the C1 subpopulation, $16.1 \pm 0.1 \%$ of cells contained a double chromosome content and $82.5 \pm 1.0 \%$ more than double.

To investigate whether subpopulations with different DNA content show physiological differences as well, we sorted the cell population at three growth rates (0.1 h^{-1} , 0.2 h^{-1} and 0.7 h^{-1}) into subpopulations containing single (C1), double (C2) or more than double chromosome content (Cx)

aiming to analyze their proteome profile as the basis of their phenotype. In total, 677 unique proteins could be detected. 351 proteins were found in at least one replicate of all subpopulations and 245 proteins were found across all replicates. 707 different functions of 647 unique proteins were annotated using the database of clusters of orthologous groups (COG) (Tatusov et al., 1997) (see Figure A.7 in section A.6 ‘Supplementary material’). 95.2 % of the control population (CP) proteome could be found in the reference population (RP) proteome without significant changes, indicating only a small influence of cell sorting on protein recovery and confirming the quality of the analysis.

Significant changes in protein quantity were defined by exceeding a threshold of more than 1.5 fold change (FC) in combination with a p -value < 0.05 (Student’s t -test). Changes in metabolic pathways were detected using GAGE and GlobalTest gene set analysis (Luo et al., 2009; Goeman et al., 2004) applying the same significance filter as for the individual proteins. As a result, at any given growth rate, the proteomic patterns of the subpopulations did not differ significantly from each other (Figure A.4a). When looking at single proteins, only three were detected that comprised significantly different levels between subpopulations at growth rate $\mu = 0.1 \text{ h}^{-1}$ and $\mu = 0.7 \text{ h}^{-1}$, respectively. The abundance of cell division protein FtsZ was found to be 3.6 fold lower in subpopulation C1 in contrast to C2. FtsZ is a bacterial tubulin homologue self-assembling into a ring at mid-cell level and localizing the bacterial divisome machinery (Adams et al., 2009; Weart et al., 2007). The two other proteins were the molecular chaperone GroEL (FC 1.7) and a P-47-like protein (PP_2007, FC 2.4). Also at high growth rate of $\mu = 0.7 \text{ h}^{-1}$, only three proteins, the translocation protein TolB (FC 1.8), the NADH dehydrogenase subunit G (PP_4124, FC 1.51) and a succinyldiaminopimelate transaminase (PP_1588, FC 0.26) showed significant differences between the subpopulations C2 and Cx. Surprisingly, no changes in metabolic pathways could be found between subpopulations at any given growth rate.

Comparing the subpopulations of different growth rates with RP, biologically significant differences were detectable as tested by gene set analysis (GAGE (Luo et al., 2009)) and Globaltest (Goeman et al., 2004)) (Figure A.4b and c). At $\mu = 0.1 \text{ h}^{-1}$, subpopulations C1 and C2 showed higher abundance of proteins related to ‘cell motility’, and proteins involved in ‘cell cycle control, cell division and chromosome partitioning’ (cell cycle) were additionally highly abundant in subpopulation C2. Apart from COG annotated pathways, several proteins connected to carbon storage were found to be significantly changed (Figure A.5). Mirroring low q_s at slow growth compared to moderate growth, four main signaling proteins in chemotaxis (CheA, CheB, CheW, CheV) as well as 6 methyl accepting chemotaxis transducers were significantly increased. Furthermore, the low abundance of glycogen synthesis proteins (GlgA, Pgm) and the high abundance of glycogen hydrolysis proteins (GlgX, GlgP) could be seen together with an increase of proteins involved in PHA production (PhaA, PhaC).

In contrast, subpopulations C2 and Cx of fast growing cells ($\mu = 0.7 \text{ h}^{-1}$) revealed higher presence

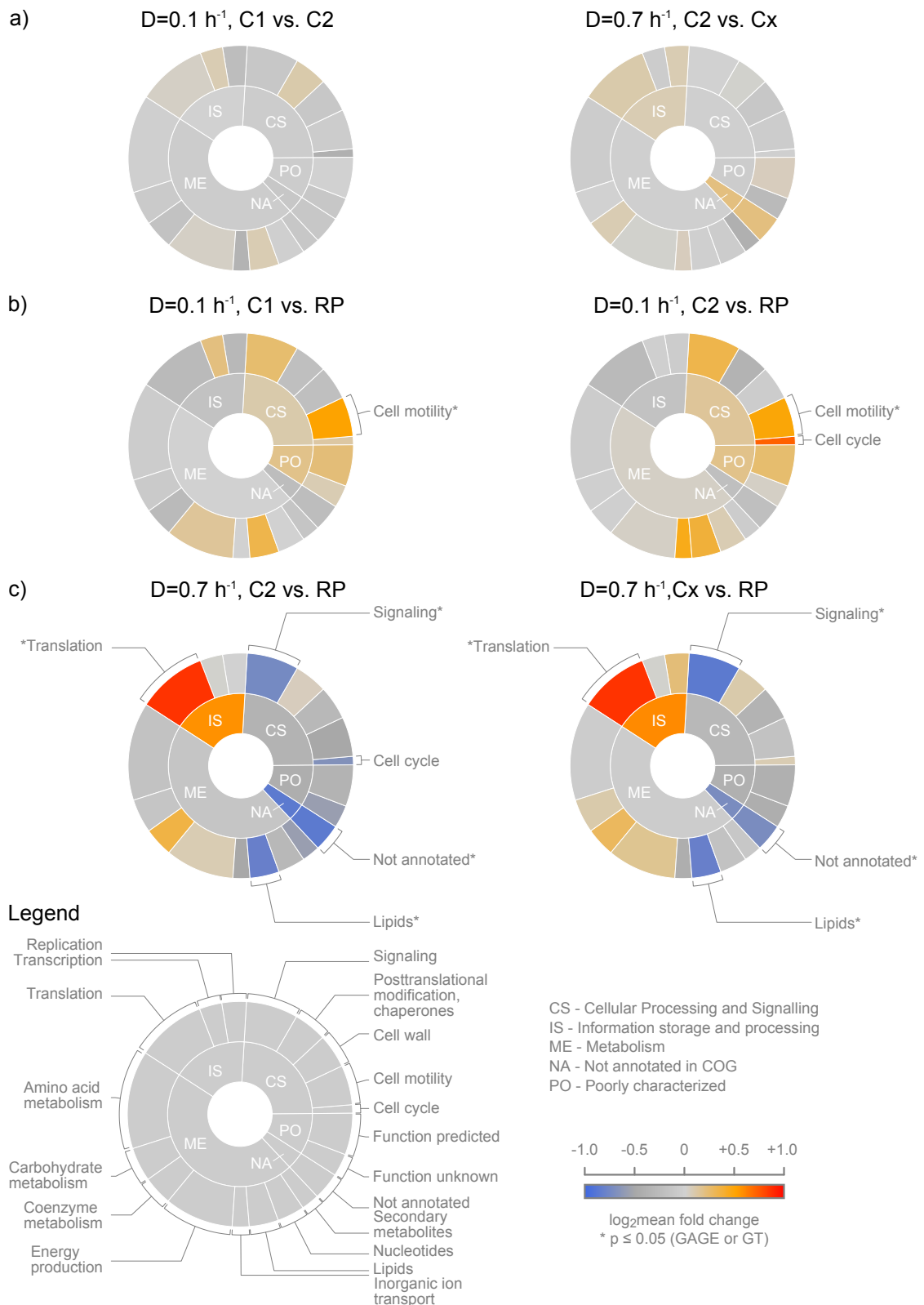


Figure A.4.: Circular treemaps visualizing differentially expressed functional protein categories. Proteins detected by mass spectrometry were clustered according to their pathway annotation in COG covering two levels of specificity (Tatusov et al., 1997). The size of a sector is proportional to the number of proteins found in one specific pathway in relation to the total protein number. The color code represents the \log_2 mean fold change ($\log_2 \text{FC}$) of protein quantity in one pathway. The color blue codes for an underrepresentation, red for an overrepresentation of the proteins in a pathway compared to the reference population (RP, $\mu = 0.2 \text{ h}^{-1}$). Pathways with a fold change in the range $\log_2 \text{FC} < -0.58$ and $\log_2 \text{FC} > 0.58$ are labeled with the respective pathway name. Pathways that were significantly changed using GAGE (Luo et al., 2009) and Globaltest (Goeman et al., 2004) gene set analysis are additionally marked (*). A. Comparison of the subpopulations C1/C2 and C2/Cx at growth rates 0.1 h^{-1} and 0.7 h^{-1} . B. Comparison of the subpopulations C1 and C2 at $\mu = 0.1 \text{ h}^{-1}$ with RP. C. Comparison of the subpopulations C2 and Cx at $\mu = 0.7 \text{ h}^{-1}$ with RP.

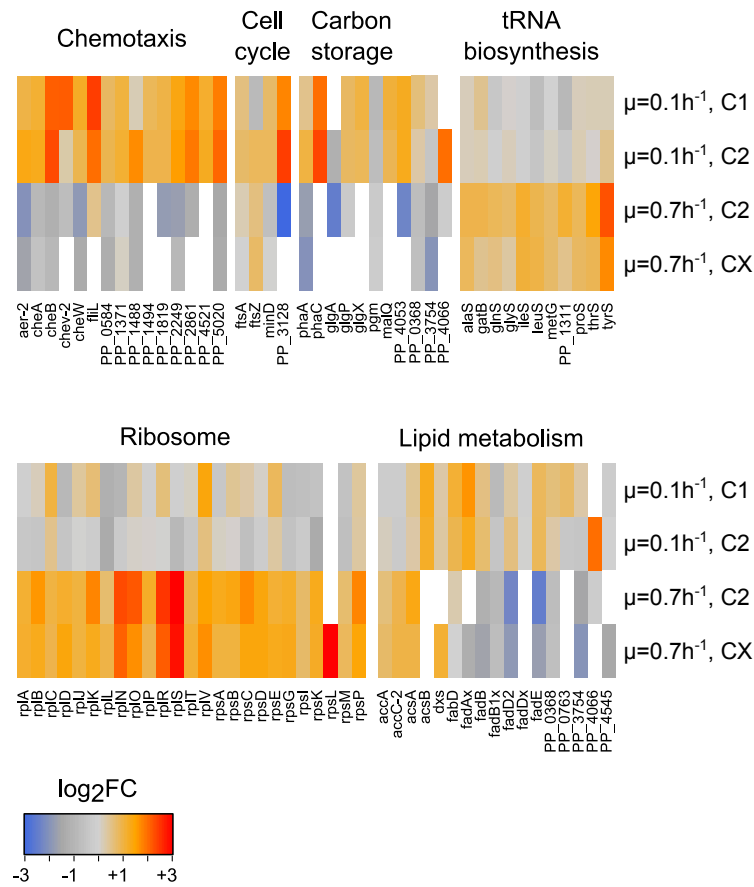


Figure A.5.: Heatmaps of metabolic pathways of special interest. The log₂-fold changes of annotated proteins are visualized ranging from blue (low abundance) to red (high abundance). A detailed annotation of the protein names can be found in the supplementary material, additional file 1. One line of the heatmap represents the different subpopulations (C1, C2 and Cx) at different growth rates ($\mu = 0.1\text{h}^{-1}$, $\mu = 0.7\text{h}^{-1}$). Proteins of the specific pathways are shown column-wise.

of proteins grouped in the pathway ‘Translation, ribosomal structure and biogenesis’ (Translation), while proteins of ‘Signal transduction mechanisms’ (Signaling) and ‘Lipid transport and metabolism’ (Lipids), were significantly underrepresented. The faster growth was reflected in proteins related to translation and therefore protein production. Here, 11 tRNA synthetases and 25 ribosomal proteins showed significantly higher abundance. In lipid metabolism, mostly enzymes of beta-oxidation were found in lower presence at fast growth (Figure A.5). The supposed down regulation of the ‘Cell Cycle’ (C2 versus Cx) was mainly due to the single protein change of the poorly characterized PP_3128.

In summary, the proteome of cells differing in DNA content but of identical growth rate was highly similar, whereas the proteome of cells cultivated at different growth rates was significantly diverging in particular pathways.

A.4. Discussion

Considering the influence of different growth rates on the population, proteome analysis revealed that slow growth triggered starvation response, while fast growing cells revealed accelerated protein synthesis and alleviated stress physiology. In slowly growing cells, proteins connected to PHA synthesis and glycerol hydrolysis were amplified, indicating higher PHA carbon storage activity. Additionally, these cells showed protein patterns anticipating increased motility and chemotaxis response. Notably, low q_S values of slowly growing cells ($\mu = 0.1 \text{ h}^{-1}$) were not reflected on the energetic state of the population. AEC values did not differ significantly between slow and moderate growth rates of 0.1 h^{-1} and 0.4 h^{-1} , respectively. Chemotaxis and cellular motility as a response to carbon-poor conditions are well-known phenomena in natural environments (Harshey, 2003; Soutourina et al., 2003). Our observations in slowly growing cells are in agreement with findings of transcriptome studies in ‘average populations’ of other species. For instance, studies in *E. coli* showed higher expression of genes involved in motility at slower growth rates in direct comparison to faster growth conditions (Nahku et al., 2010) and studies in *Saccharomyces cerevisiae* showed significant amplification of carbon storage metabolism at slow growth (François et al., 2001).

Fast growing cells were obviously investing resources in proteins involved or related to the translation machinery. Multiple ribosomal proteins as well as tRNA synthetases were highly abundant fostering protein/biomass production (Figure A.5). This finding is also in agreement with observations in eukaryotes like *S. cerevisiae* (Rebnegger et al., 2014) and prokaryotes such as *Salmonella typhimurium* (Schaechter et al., 1958). Additionally, proteins of typical carbon storage pathways e.g. PHA synthesis were less abundant in *P. putida* KT2440. Proteins of lipid biosynthesis, especially involved in beta oxidation were also lowered in fast growing cells compared to RP. This observation is in agreement with the lower abundance of the PHA synthesis proteins, as the beta oxidation provides precursors (Aldor et al., 2003).

To our surprise, the almost 6.5-fold increase of the specific glucose uptake rate with increasing growth rate (Figure A.2), was not mirrored by major changes among proteins involved in carbohydrate and energy metabolism.

Notably, relative changes of protein quantity can be elucidated with the method applied here. Absolute changes per cell, dependent on the growth rate were not measured with the applied workflow, as it was first shown for the sum of proteins by Schaechter et al. (1958). Their pioneering studies described an exponential increase in protein, DNA and RNA contents and therefore, cell size with increasing growth rates (Bremer et al., 2004; Maaløe et al., 1966; Schaechter et al., 1958). In our study, the relative cell size estimation was acquired using FCS. In accordance to various other cell cycle analyses, the FCS increased with increasing growth rates (Donachie, 1968; Hewitt et al., 1999; Skarstad et al., 1983; Neumeyer et al., 2013) (Figure A.3). Following the rationale of Schaechter et al. (1958), this phenomenon reflects increasing protein contents per cell. We presume

that the increased amount of cellular glucose uptake is proportional to the elevated production of proteins, thus increasing absolute protein quantity but leaving relative quantity unchanged.

Studying the putative impact of growth rate and cell cycle stage on the functional diversity of a population, the growth rate is obviously a major determinant for cellular protein composition, as found in our chemostat studies. Growth and cell cycle were clearly linked, but subpopulations showing different DNA content showed only small differences in cellular physiology at the same growth rate. The detection of FtsZ in a significant higher abundance in the C2 subpopulation, which is preparing for division after finishing replication, is in agreement with its assigned function as a proposed diffusible factor (Teather et al., 1974) initiating cell division (Chien et al., 2012). Despite this cell cycle related finding, subpopulations showed almost identical protein patterns irrespective of cell sizes, anticipated protein mass (Lindmo, 1982; Rønning et al., 1979) and DNA content.

Surprisingly, no signs for a specialization of cells in different cell stages for e.g. carbon storage or protein production/growth could be observed that could support the hypothesis of shared tasks of subpopulations in B- and pre-D/D-phases during the cell cycle. This result is remarkable: subpopulations distinguished by DNA content appear to be physiologically highly similar provided that the growth rate is the same.

Although we are aware that subpopulations do not mirror single cell proteome compositions the high resemblance of the subpopulations proteome patterns at the various growth rates point to their nearly identical physiological state.

One may argue whether this finding was influenced by the operation mode ‘chemostat’. We identified the high similarity among subpopulations by installing distinct growth rates, because superimposing impacts in classical (fed-) batch fermentations would have prevented the unequivocal growth-to-subpopulation analysis. However, the chemostat approach might have excluded the detection of subpopulations with different protein contents because this ‘growth rate filter’ was installed. Assuming that cells aim to grow with the least energetic burden as possible, cellular protein compositions should be optimized at a given growth rate. Therefore, it could not be excluded, that subpopulations showing different protein patterns may have existed, but were washed-out because they could not achieve the required growth rate. While the latter demands for further in-depth analysis, the determining impact of growth on cell cycle and subpopulations is clearly visible. It gives rise to the assumption that the cell cycle itself has a minor impact on population heterogeneity under the conditions tested.

A.5. Acknowledgments

This work was supported within the ERA-IB / ERA-NET scheme of the 6th EU Framework Programme (0315932B).

A.6. Supplemental material

Additional file 1

Additional file 1 contains the dataset of the label-free quantification (LFQ) values that were used for further analysis of differences in protein pattern. The file can be downloaded on '<http://www.amb-express.com/content/4/1/71/additional>'.

Additional file 2

Additional file 2 contains the supplementary Figure A.6 and Figure A.7.

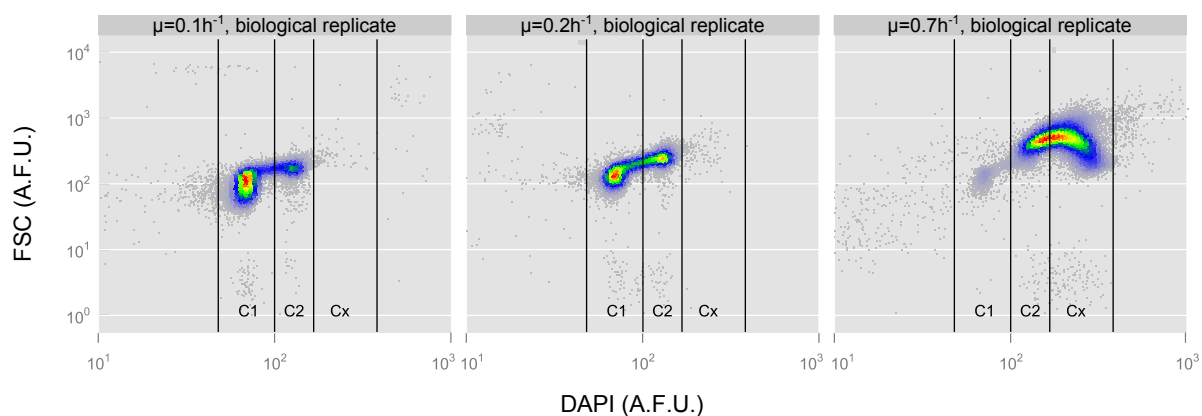


Figure A.6.: Replicate dataset of dot plots of DNA content (DAPI, in arbitrary fluorescence units (A.F.U.)) versus forward scatter (FSC, in A.F.U.) at different growth rates 0.1h^{-1} , 0.2h^{-1} and 0.7h^{-1} . The DNA content and the forward scatter increased with increasing growth rate. The indicated gates (C1, C2, Cx) were used for sorting 5×10^6 cells per subpopulation for further mass spectrometric analysis.

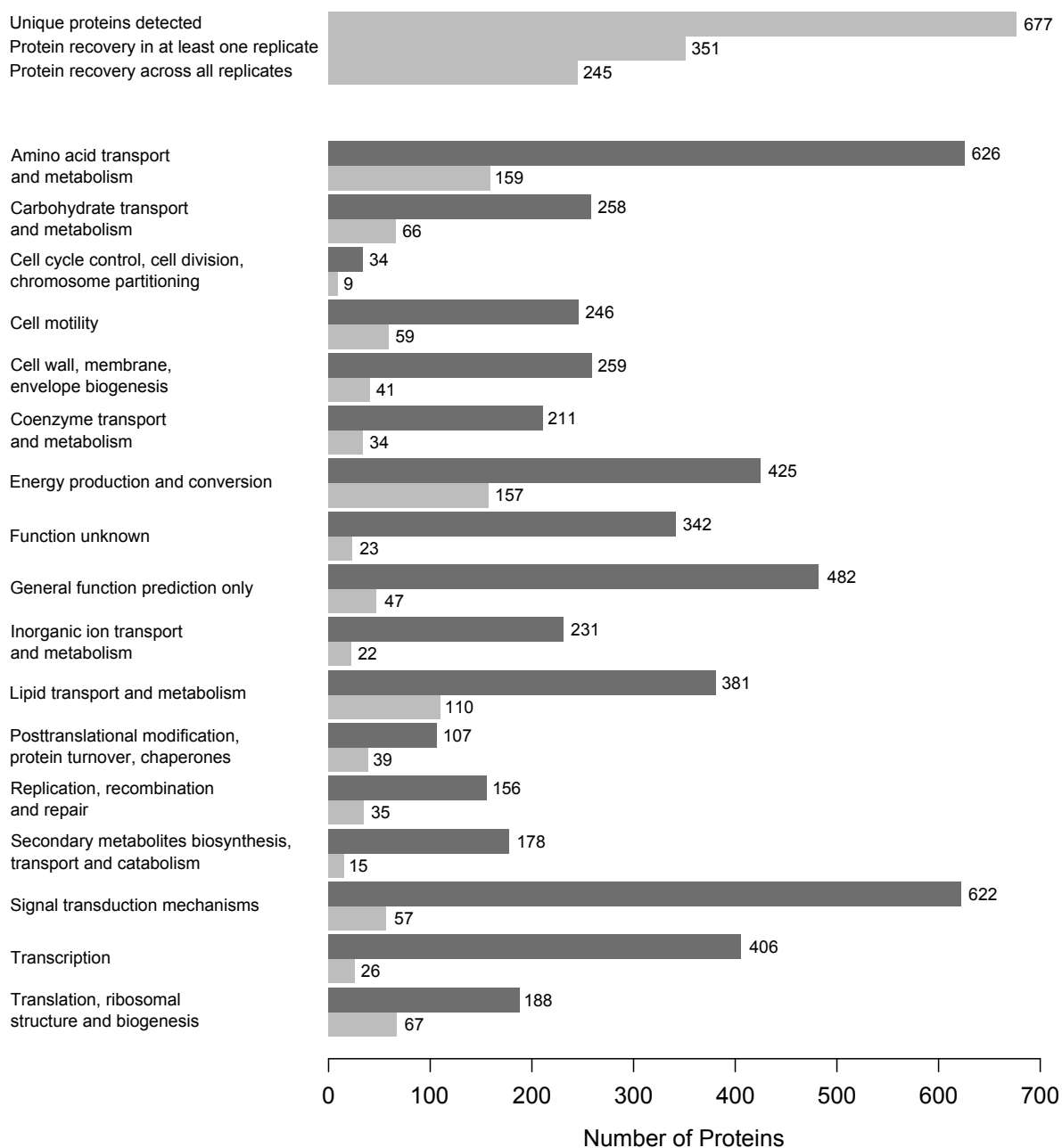


Figure A.7.: Overview of the total protein detection and protein annotation. Overall, 677 unique proteins were identified, 351 proteins were detected in at least one replicate of all subpopulations and 245 proteins were found across all replicates. Functional annotation was carried out using the COG database (Tatusov et al., 1997). 707 different functions of 647 unique proteins could be annotated into 17 categories. The total number of proteins of *Pseudomonas putida* KT2440 annotated in one specific category (dark grey bars) is compared to the number of proteins recovered in this study (light grey bars).

B. Manuscript II

Cellular response to different types of stress is the hallmark of the cell's strategy for survival. How organisms adjust their cell cycle dynamics to compensate for changes in environmental conditions is an important unanswered question in bacterial physiology. A cell using binary fission for reproduction passes through three stages during its cell cycle: a stage from cell birth to initiation of replication (B phase), a DNA replication phase (C phase) and a period of cell division (D phase).

We present a detailed analysis of durations of B, C, and D phases, investigating the cell cycle dynamics under environmental stress conditions. Applying continuous steady state cultivations (chemostats), the DNA content of a *Pseudomonas putida* KT2440 cell population was quantified with flow cytometry at distinct growth rates. Data-driven modeling revealed that the maximum replication rate of *P. putida* KT2440 is similar to *Escherichia coli* and to other organisms using symmetric binary fission for reproduction. Under stress conditions, such as oxygen deprivation, solvent exposure and decreased iron availability, DNA replication was accelerated significantly, correlated to the severity of the imposed stress (up to 1.9 fold). Transcriptome data underpin the transcriptional upregulation of crucial genes of the replication machinery to achieve the replication speed up.

We show that fast replication of the genetic information is of high priority under stress conditions and that a balanced altering of the duration of cell cycle phases is a cellular strategy to maintain constant growth rates under stress.

B.1. Introduction

Binary fission represents one of the most common ways of reproduction within the domain of bacteria (Chien et al., 2012). It is dominated by two major mechanisms: DNA replication and cell division. In terms of cell cycling, their order is classically represented by a three-sectional circuit, consisting of a B, C and D period: The B period, which is defined as the time between cell birth and initiation of replication, the C period in which the chromosome is replicated and the D period representing the remaining time between termination of replication and end of cell division.

In their pioneering studies, Cooper and Helmstetter (1968) succeeded to model *Escherichia coli*'s cell cycle. Their mathematical approach implemented two fundamental rules: The cell does not start replicating its DNA unless a critical threshold is achieved and it does not divide unless two genomes are present. Additionally, the concept of multifork replication was included.

This chapter has been published as: **Sarah Lieder, Michael Jahn, Joachim Koepff, Susann Müller and Ralf Takors** (2015) Environmental stress speeds up DNA replication in *Pseudomonas putida* in chemostat cultivations. *Biotechnology Journal* 11(1):155-63

This model correlates the individual lengths of B, C and D periods with cell growth. The authors found nearly constant C and D periods for well growing cells while the B period diminished with increasing growth rate. The duration of the B period is coupled to a constant critical cell mass in *E. coli* (Donachie, 1968). This critical cell mass is either already present or rapidly reached by the cell under nutrient-rich conditions, while more time is needed in nutrient poor media. Further, Helmstetter (1996) illustrated that C periods are longest at slow growth and decreased steadily to constant values under alleviated growth conditions. The D period is also described to be relatively constant and even determining the generation time when replication is uncoupled (Cooper et al., 1968). Müller extended this model by introducing the pre-D phase that occurs under limiting, even harsh growth conditions (Müller, 2007). The pre-D period specifies the bacterial disability to divide after finishing replication, obviously waiting for improved growth conditions. Consequently, the pre-D period disappears under optimal growth conditions similar to the B phase.

Summarizing, the work of Cooper and Helmstetter (1968) provided a mathematical model correlating successfully the basics of DNA replication and cell division. While their approach of modeling the cell cycle as a single process could be well applied for *E. coli*, one may argue whether the inherent link of DNA replication (C phase) and cell division (D phase) is too general, and the cell cycle consists of coordinated but independent processes instead (Wang et al., 2009).

So far, experimental studies followed the classical motivation of investigating nutrient-rich versus nutrient-poor growth conditions, thus correlating durations of B, C and D periods exclusively with velocity of cell growth. The D period was already described to be prolonged under stressful conditions (pre-D phase) (Müller, 2007). In this study, we argue that the duration of the C phase might as well be not as strictly connected to the growth rate as presumed in the past, but that an independent adjustment of cell cycle phase durations might be a possible strategy to survive stressful conditions.

Since many years, flow cytometry (FC) has proven to be an excellent and powerful tool for the investigation of the cell cycle in a precise, robust and high throughput way by stoichiometric fluorescent labelling of DNA (Müller et al., 2010; Srienc, 1999; Steen, 2001). The number of subpopulations and the number of individuals comprising the subpopulations create a characteristic pattern, providing insights into the duration of cell cycle phases when combined with the cell cycle model of Cooper and Helmstetter (Cooper et al., 1968; Skarstad et al., 1985; Cooper, 1991).

This study deals with the question whether and how stress imposed on bacteria affects the interplay of DNA replication and cell division mirrored by the duration of B, C and D periods. In contrast to commonly used nutrient-rich/-poor experiments, we performed carbon limited chemostat cultivation using *P. putida* KT2440, additionally applying stress conditions, such as limited oxygen supply, organic solvent addition (5% v/v decanol) or decreased iron availability. The appearance of distinct subpopulations in DNA content was analyzed via FC at a given growth rate. Notably,

these continuous steady state conditions prevent the overlay of different cell states usually occurring in batch experiments (Skarstad et al., 1985; Wiacek et al., 2006). Instead, chemostats select for equally fast growing cells, even when stress conditions are applied.

B.2. Material and Methods

Growth conditions

Chemicals were purchased from Fluka, St. Gallen, Switzerland. Experiments were performed with cells originating from a single colony stored in a working cell bank at -70°C . Cells were cultivated in M12 minimal salt medium containing 2.2 g L^{-1} $(\text{NH}_4)_2\text{SO}_4$, 0.4 g L^{-1} $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 0.04 g L^{-1} $\text{CaCl}_2 \cdot \text{H}_2\text{O}$, 0.02 g L^{-1} NaCl , 2 g L^{-1} KH_2PO_4 and trace elements (2 mg L^{-1} $\text{ZnSO}_4 \cdot \text{H}_2\text{O}$, 1 mg L^{-1} $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 15 mg L^{-1} $\text{Na}_3\text{-citrate} \cdot 2\text{H}_2\text{O}$, 1 mg L^{-1} $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 0.02 mg L^{-1} $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$, 0.03 mg L^{-1} $\text{NaMoO}_4 \cdot 2\text{H}_2\text{O}$, 0.3 mg L^{-1} H_3BO_3 , 10 mg L^{-1} $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$). The carbon source glucose was supplied at a concentration of 5 g L^{-1} and 10 g L^{-1} in shake flask and bioreactor cultivations, respectively.

Bioreactor cultivations were inoculated with a 150 mL mid-exponential shake flask pre-culture (1 L baffled shake flask). The inoculum was transferred into a bioreactor (KLF 3.7 L, Ser. No. 10819, Bioengineering AG, Wald, Switzerland) to reach a final working volume of 1.5 L. The environmental conditions were set previous to inoculation to 30°C , a stirrer speed of 700 rpm, a pressure of 0.5 bar and an aeration of 2 L min^{-1} sterile filtered ambient air. The pH was set and maintained at pH 7 with 25% (v/v) NH_4OH . Exhaust gas composition (Blue Sense CO_2 and O_2 , (DCP- CO_2 DCP- O_2 , Blue Sense gas sensor GmbH, Herten, Germany), dissolved oxygen and pH in the liquid phase (Ingold, Mettler Toledo GmbH, Giessen, Germany) were monitored online. The batch phase was continued as chemostat when glucose was depleted.

The dilution rate was controlled by weight gain of the bioreactor: A medium feed of 10 g was the control variable for the harvest pump to remove 10 g of biosuspension. The dilution rate, and therefore, the growth rate, was crosschecked manually by measuring the mass of the harvest outflow within a timespan of one hour before sampling. Steady state was evaluated online via exhaust air data. Detailed information about the theoretical background of a chemostat can be found in the supplemental file S1 in section B.6.

For *standard cultivations*, the growth rate was stepwise increased from $\mu = 0.1\text{ h}^{-1}$ to $\mu = 0.7\text{ h}^{-1}$ until a clear wash-out of cells could be detected.

For *stress investigations*, a constant growth rate of $\mu = 0.2\text{ h}^{-1}$ was maintained throughout the cultivation. At first, the culture was grown under reference conditions (all nutrients were supplied in excess, except glucose). After five residence times, the environmental condition was changed to the stress condition and kept constant for additional five residence times. Finally, the culture was

shifted back to reference conditions, to make sure, that cells revert to their original physiological condition, and that the population composition was not changed e.g. by putative selection of mutated strains. The stress conditions included decreased iron availability (50 % reduction of the iron source in the media composition), oxygen deprivation ($pO_2=5\%$ and $pO_2=1.5\%$; A $pO_2=100\%$ was defined as the dissolved O_2 level in the bioreactor under operating conditions, but without biomass in suspension) and solvent exposure (5 % v/v decanol).

Analytics

Residual glucose concentrations in the supernatant of the biosuspension were quantified via a D-glucose measurement kit according to the manufacturer's instructions (R biopharm AG, Darmstadt, Germany). The cell dry weight (CDW) was measured for mass based calculations by taking a suspension sample of 40 mL; 10 mL each were filled into preliminary weighed glass tubes, centrifuged at $5,500 \times g$ and $4^\circ C$ for 10 min and washed twice with 5 mL 0.9 % w/v NaCl. The pellet was dried for 48 h in an $85^\circ C$ chamber before the mass gain of the glass tubes were measured.

Flow Cytometry

1 mL of cultivation broth was taken directly into precooled 0.9 % w/v NaCl solution, centrifuged for 5 min at $5,000 \times g$ at $4^\circ C$, washed with PBS, resuspended in cryo-protective solution (15 % glycerol in PBS according to Jahn et al. (2013) and stored at $-20^\circ C$.

Samples were thawed on ice and centrifuged for 2 min at $8,000 \times g$ and $4^\circ C$ to remove cryo-protective solution. The supernatant was discarded, the cells were resuspended in ice cold PBS and adjusted to an optical density of $OD_{600nm}=0.05$ in 2 mL volume. For DNA staining the cells were harvested by centrifugation, taken up in 1 mL permeabilization buffer (0.1 M citric acid, 5 g L^{-1} Tween 20), incubated for 10 min on ice and harvested by centrifugation. Finally, cells were resuspended in 2 mL ice cold staining buffer (0.68 μM DAPI, 0.1 M Na_2HPO_4), filtered through a Partec CellTrics mesh with 30 μm pore size and stored on ice until analysis.

Flow cytometry was performed using a MoFlo cell sorter (Beckman-Coulter, USA) as described before (Jehmlich et al., 2010). The DAPI fluorescence was recorded using a multi-line UV laser for excitation (333-365 nm, 100 mW) and a bandwidth filter for emission (450 ± 30 nm). The datasets were annotated according to the miFlowCyt standard (Lee et al., 2008) and are publicly available on the FlowRepository database (Spidlen et al., 2012).

Mathematical modeling

The duration of the cell cycle phases C and D' were calculated iteratively, minimizing the distance of the theoretical DNA content (calculated according to the mathematical model of Cooper and Helmstetter (1968)), to the flow cytometrically measured DNA distributions $n(G)_{\text{exp}}$ (Skarstad et al., 1985). A detailed description of the implementation of the mathematical model can be found in the supplemental file S2 in section B.6.

Transcriptome Analysis

Sampling procedure A sample of 2 mL cultivation broth was taken directly into 4 mL of RNAprotect Bacteria Reagent (Qiagen GmbH, Germany), vortexed and incubated at room temperature for 5 min. Aliquots of the solution containing approximately 10^9 cells were centrifuged at 7000 x *g* for 10 min at 4 °C. The supernatant was discarded and the cell pellet was shock frozen in liquid nitrogen and stored at -70 °C.

RNA next generation sequencing The samples were collectively shipped on dry ice for a batch RNA next generation sequencing, carried out by MFT Services (Tübingen, Germany). Ribosomal RNA species were removed from the sample RNA using the RiboZero rRNA Removal Kit (Epicenter). Sequencing libraries were prepared with the TruSeq™ RNA Sample Preparation Kit v2 (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's instruction and quantified with a Qubit® fluorometer (Life Technologies, Carlsbad, USA). Equimolar amounts were loaded onto an Illumina GAIIx flow cell (Illumina, Inc., San Diego, CA). Bound molecules were clonally amplified on a cBot instrument (Illumina, Inc., San Diego, CA). The quality controlled (Andrews, 2010) fastq sequences were aligned against the *P. putida* KT2440 genome (AE015451.1) using bowtie v0.12.7 (Langmead et al., 2009). Reads mapping to rRNA loci were removed before the quantification step. HTSeq (Anders, 2010) was used to count reads. The statistical data analysis was performed with the bioconductor package 'edgeR' (Robinson et al., 2010). Raw count data were first normalized based on 'counts per million mapped counts' (CPM), to account for differences in sequencing depth. Discrete count data as obtained by RNA-Seq was shown to follow a negative binomial (NB) distribution (McCarthy et al., 2012). Differential expression analysis was carried out following the protocol by Anders et al. (2013) using edgeR (Robinson et al., 2010). *p*-values were adjusted for multiple testing according to Benjamini and Hochberg (1995) to calculate the false discovery rate (FDR). A cutoff of $\text{FDR} \leq 0.05$ was chosen to extract differentially expressed genes. Genes were categorized into functional groups using COG (clusters of orthologous groups) (Tatusov et al., 1997).

B.3. Results

Two different experimental approaches were combined with data-based mathematical modeling to investigate stress related influences on cell cycle dynamics (Figure B.1a and b).

Standard experiments were performed under carbon limited conditions in chemostats. The growth rate was stepwise increased until the maximum growth rate of *P. putida* KT2440 ($\mu = 0.7 \text{ h}^{-1}$) was reached, resulting in the wash-out of the population (Figure B.1c). The experimental condition at growth rate $\mu = 0.2 \text{ h}^{-1}$ is referred to as reference condition.

Stress experiments were performed at a constant growth rate of $\mu = 0.2 \text{ h}^{-1}$. The stress-shift was introduced and kept until cells had adapted to the new conditions showing steady state growth, notably at the same growth rate of $\mu = 0.2 \text{ h}^{-1}$. Afterwards, the culture was shifted back to reference conditions (Figure B.1e). Comparing the population before and after the stress exposure, identical physiological features were observed (carbon emission rate, cell dry weight) and the population composition did not change regarding the parameters monitored by flow cytometry.

Cell cycle analysis of standard steady state cultures

Flow cytometry revealed characteristic subpopulations with different chromosome contents for each growth rate (Figure B.1d). Four different subgroups of cells could be identified and allocated to the specific cell cycle phases:

- A subpopulation with a single chromosome content representing cells in B phase that just divided and did not start replication yet (subB)
- A subpopulation with a chromosome content between single and double, representing cells in replication phase C (subC)
- A subpopulation containing the double chromosome content in pre-D or D phase, representing cells that finished replication but did not divide yet (subD'). As it is not possible to distinguish between pre-D and D phase by DNA content, these two phases were merged and were referred to as D' phase
- A subpopulation was found at high growth rates with more than doubled chromosome content representing cells performing multifork DNA replication (subMF)

The analysis of DNA content of cells with slow to moderate growth (0.1 h^{-1} - 0.4 h^{-1}) showed that fractions of subB decreased with increasing growth rate while subD' increased. The growth rate of $\mu = 0.4 \text{ h}^{-1}$ can be qualified as an inherent threshold in *P. putida* KT2440, as cells started to uncouple DNA replication from cell division with increasing growth: At higher growth rates than $\mu = 0.4 \text{ h}^{-1}$, no more cells were present in B phase (subB), and an increasing fraction of subMF cells with uncoupled cell cycle and a decreasing portion of subD' was observed.

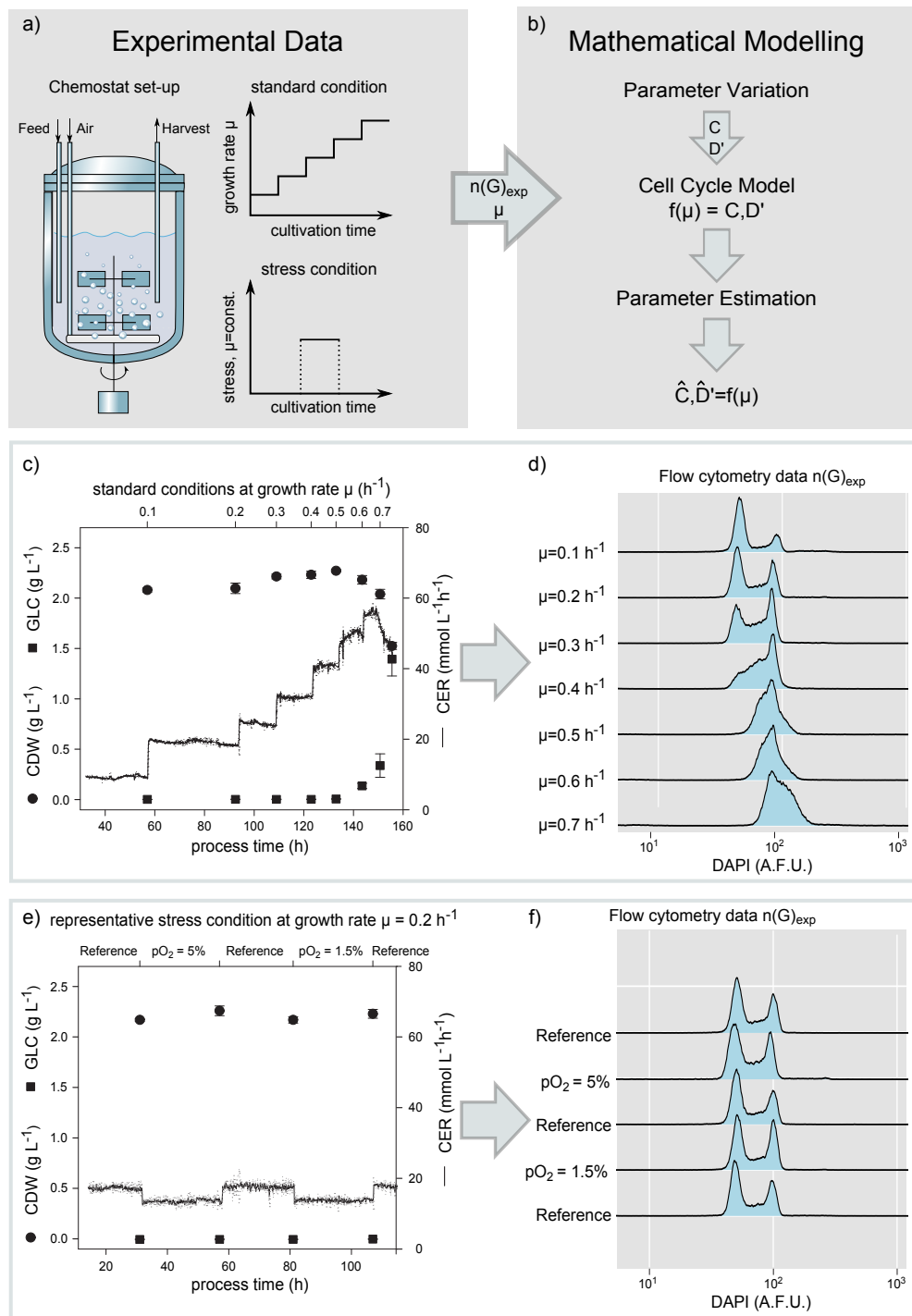


Figure B.1.: Overview of the experimental set-up (a) and the workflow of data-based modeling (b). Chemostats were carried out under standard and stress conditions. At standard conditions, all nutrients except glucose were supplied in excess and the growth rate was stepwise increased until wash-out occurred. Three different stress conditions were applied in a shift like manner: decreased iron availability, deprivation of oxygen and solvent exposure. The model parameters, duration of C and D' phase, were fitted by non-linear regression using the flow cytometry data ($n(G)_{\text{exp}}$) and the growth rate μ . Physiological data at standard and at a representative stress condition (oxygen deprivation) are shown in c) and e), respectively. Biomass concentrations (CDW, black dots, g L^{-1}) and residual glucose concentrations (GLC, black squares, g L^{-1}) were measured after 5 residence times of one specific dilution rate at steady state. The carbon dioxide emission rate (CER, black line, $\text{mmol L}^{-1} \text{h}^{-1}$) was monitored online. The error bars and lines represent the standard deviation of biological triplicates. A summary of the flow cytometry data is given in d) and f). DNA histograms (DAPI, arbitrary fluorescence units A.F.U.) at different growth rates (d) and different environmental conditions (f) are depicted.

Table B.1.: Summary of the duration of cell cycle phases and goodness of fit of the simulation. Average values of calculated \hat{B} , \hat{C} and \hat{D}' phases (h) of 3 biological replicates and their standard deviation were calculated on the basis of the mathematical model of Cooper and Helmstetter (1968). s is the deviation of the simulated to the experimental number of cells measured by flow cytometry and presented as subpopulation distributions in DNA histograms. The formula was framed by Skarstad et al. (1985)

growth rate μ	\hat{C} (h)	\hat{D}' (h)	\hat{B} (h)	s
0.1 h^{-1}	3.48 ± 0.01	1.01 ± 0.17	2.41	0.55 ± 0.1
0.2 h^{-1}	1.54 ± 0.04	0.94 ± 0.07	0.92	0.69 ± 0.35
0.3 h^{-1}	1.38 ± 0.04	0.63 ± 0.06	0.29	0.49 ± 0.14
0.4 h^{-1}	1.20 ± 0.03	0.59 ± 0.05	0	0.40 ± 0.18
0.5 h^{-1}	1.04 ± 0.02	0.66 ± 0.01	0	0.38 ± 0.21
0.6 h^{-1}	1.03 ± 0.02	0.57 ± 0.04	0	0.84 ± 0.26

Using the implemented mathematical model (supplemental file S2 in section B.6), the durations of the cell cycle phases C and D' were calculated (Table B.1). Notably, the standard deviation was less than 5 %, underpinning the chemostat approach as a reliable and reproducible tool for the investigation of cell cycle phases.

At standard conditions, the duration of replication phase C was decreasing with increasing growth rate until a minimal length of $C_{\min} = 62 \text{ min}$ was reached (Figure B.2). Strikingly, comparing the calculated replication phase durations of *P. putida* KT2440 with previous results of *E. coli* B/r strains (Helmstetter, 1996) very similar trajectories and replication times were found. To evaluate the similarity of *E. coli* and *P. putida* replication rates (r_c), the exponential model of Keasling et al. (1995) was applied, which describes the dependency of C phase duration and growth rate in *E. coli*. A reasonably high goodness of fit ($R^2=0.95$) was found with the pooled data of *E. coli* and *P. putida*, supporting the high similarity of the results between the two organisms. Combining the minimum replication time $C_{\min} = 62 \text{ min}$ with the chromosome size of *P. putida* KT2440 (6.18 Mb) (Nelson et al., 2002), a maximum replication rate of $r_c \approx 100 \text{ kbp/min}$ could be calculated at standard conditions.

Cell cycle analysis of stressed steady state cultures

To investigate the impact of stress on cell cycle kinetics, chemostat cultivations were performed at $\mu = 0.2 \text{ h}^{-1}$ with additionally imposed respiration stress (reduction of dissolved O_2 to 5 % and 1.5 % partial pressure $p\text{O}_2$), environmental stress (presence of the organic solvent decanol (5 % v/v)) and decreased iron availability. As depicted in Figure B.3, the length of the replication phase C decreased from 92 min at the reference condition to 48-64 min, depending on the severity of the stress condition. Consequently, the replication rate increased. At low oxygen partial pressure $p\text{O}_2=5 \%$, the replication rate rose 1.5-fold from 67 to around 99 kbp/min, equaling the maximal

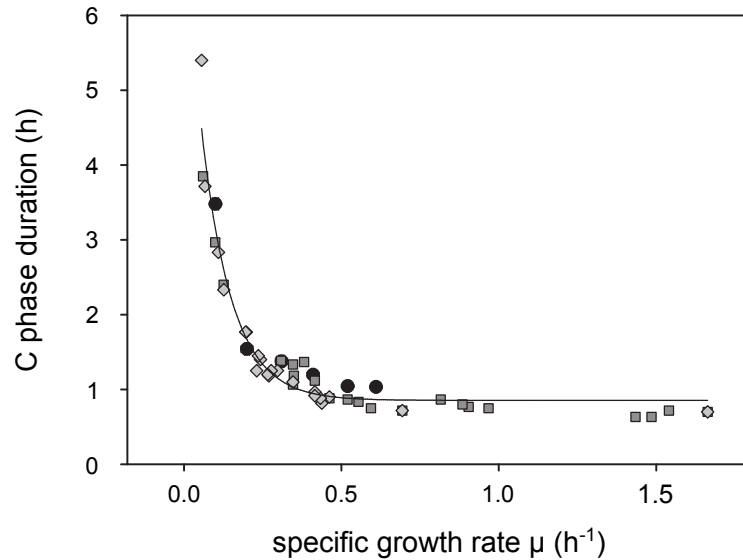


Figure B.2.: Durations of the replication phase in dependence of the specific growth rate μ . The replication time was calculated according to Cooper and Helmstetter (1968) as arithmetic mean of three biological replicates. Error bars show the calculated standard deviation. The duration of the replication (C) is decreasing with increasing growth rates until a minimum duration is reached. Black dots depict the C phase durations of *P. putida* KT2440 under steady state standard conditions. Dark grey squares (*E. coli* B/r A) and light grey diamonds (*E. coli* B/r K) show data compiled by Helmstetter et al. (1996). The pooled data could be reasonably well fitted ($R^2=0.95$) by an exponential function (black line) (Keasling et al., 1995).

replication rate found under standard conditions at $\mu = 0.7 \text{ h}^{-1}$. Surprisingly, when harsher conditions of $\text{pO}_2=1.5\%$ or decanol exposure (5% v/v) were installed, the replication rate even increased above the maximum of the standard conditions, namely 1.6-fold (110 kbp/min) and 1.9 fold (129 kbp/min), respectively. Note that all cells still showed stable steady state growth of $\mu = 0.2 \text{ h}^{-1}$, which means, that the generation time τ itself did not change but the individual contributions of B, C and D' phases varied. Apart from the shortened C phase, a clear prolongation of B and D' phases was observed (Figure B.3).

Expression profile of genes related to 'replication, recombination and repair' in stressed steady state cultures

To get a deeper insight into the mechanism of replication speed-up, the genome-wide expression profile of *P. putida* KT2440 was analyzed via next generation sequencing of the mRNA pools. Therefore, pair-wise comparison of mRNA levels between the reference condition and the most prominent stress condition -decanol exposure- was performed. Out of 5421 transcripts, which were found in total, decanol exposure caused significant changes (\log_2 fold changes (FC) > 0.58) in the expression of 540 transcripts, including 154 open reading frames with unknown function (see supplemental dataset 1 in section B.6). The 387 genes with annotated functions were categorized into functional groups using the COG database (Tatusov et al., 1997). We found 27 significantly

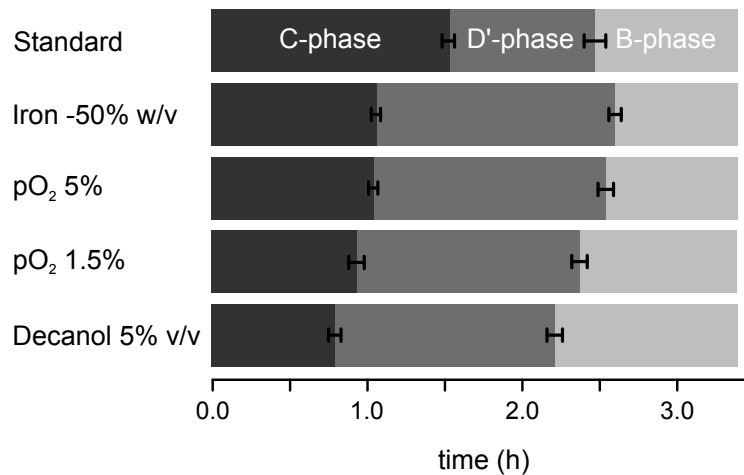


Figure B.3.: Durations of cell cycle phases in dependence of the respective stress condition, i.e. decreased iron and oxygen availability, as well as decanol exposure. The duration of the cell cycle phases is shown for all conditions tested at a growth rate of $\mu = 0.2 \text{ h}^{-1}$, corresponding to a generation time of 3.4 h. The C phase was shortened under all stress conditions in comparison to the standard condition, while B and D' phases were prolonged.

changed genes with known or anticipated tasks in ‘replication, recombination and repair’ (see supplemental dataset 1 in section B.6). Thereof, 8 genes were sorted into the functional group ‘replication’. This group showed a significant increase of expression upon decanol exposure (average \log_2 FC 0.93, Table B.2).

Among the 8 significantly changed genes, especially DNA polymerases showed elevated transcription levels. DNA ligase LigA and DNA polymerase subunits δ , χ , ϵ and τ (HolA, HolC, DnaQ and DnaX) are parts of the replication machinery. LigA catalyzes the formation of phosphodiester bonds between 5'-phosphoryl and 3'-hydroxyl groups in double-stranded DNA. It is essential for DNA replication and repair of damaged DNA. DnaQ holds the 3'-5'-proofreading exonuclease and was shown to turn the rather slow and weakly processive polymerase III core into a fast and highly processive polymerase (Studwell et al., 1990). HolA is binding the β -subunit of the DNA clamp. Johnson et al. (2005) found, that the polymerase speed is increased when the polymerase core is coupled to the β -clamp. HolC and DnaX are part of the clamp loader. HolC binds the single stranded DNA binding protein, protecting single stranded DNA and melting hairpins. DnaX connects the core polymerases to the central clamp loader and connects the replicase to the DnaB helicase. The unwinding rate of DnaB was found to be increased when bound to DNA polymerase subunit τ (Kim et al., 1996). Altogether, the most prominent transcriptional upregulation was found for genes encoding basic enzymes that are essential for a fast, efficient and processive replication.

Besides genes associated with DNA replication, 9 out of 11 significantly changed genes grouped into ‘DNA repair’ were upregulated as well under decanol stress conditions. The proteins RecB and RecD are part of a multifunctional enzyme recognizing blunt or near-blunt ends of duplex DNA,

Table B.2.: Differentially expressed genes under decanol stress conditions, annotated in the functional group ‘replication’. The COG database was used for functional annotation (Tatusov et al., 1997). The \log_2 fold change (FC) is the logarithmic ratio of expression of decanol condition and reference condition. Statistical significance was defined at a cutoff of the false discovery rate $FDR < 0.05$ (Benjamini et al., 1995).

Gene ID	Gene Product Name	$\log_2(\text{FC})$
PP_0979	DNA polymerase III subunit χ , HolC	1.26
PP_4141	DNA polymerase III subunit ϵ DnaQ	1.10
PP_4768	DNA polymerase III; subunit ϵ	1.02
PP_4796	DNA polymerase III subunit δ HolA	0.97
PP_4269	DNA polymerase III subunits γ and τ DnaX	0.94
PP_5310	ATP-dependent DNA helicase RecG	0.89
PP_4274	NAD-dependent DNA ligase LigA	0.67
PP_5088	Primosome assembly protein PriA	0.59

degrading ssDNA and dsDNA and additionally showing a DNA helicase activity (Kogoma, 1997). The RecG protein is a Holliday-junction-specific DNA helicase which is thought to catalyze reverse branch migration and was proposed to increase efficiency for homologous recombination and DNA repair (Whitby et al., 1994). The RuvC protein is a nuclease that resolves Holliday junctions in the late stages of homologous recombination (West, 1996). MutS and MutL are part of a system for recognizing and repairing errors in replication and homologous recombination. Together with the exonucleases, these enzymes are associated mainly with DNA repair and restart of replication at stalled replication forks (Kogoma, 1997).

The remaining transcripts that were affected by decanol stress were linked to the already known impact of solvent stress on the physiology of *P. putida* species, such as altered membrane composition and carbon, lipid and energy metabolism (see supplemental dataset 1 in section B.6) (Heipieper et al., 2007).

B.4. Discussion

Cell cycle kinetics were investigated combining chemostat cultivation, flow cytometry and mathematical modeling. The chemostat approach allowed us to get an unbiased insight into the effects of stress on the different phases of cell cycling at a distinct growth rate. This approach allowed high reproducibility, as mirrored by the biological variation of less than 5% between biological triplicates (Table B.1).

Applying the basic modeling assumptions of Cooper and Helmstetter (Cooper et al., 1968; Skarstad et al., 1985), cell cycle phases could be calculated successfully. We found the simulated DNA

histograms matching well to the experimentally derived DNA histograms (Table B.1, supplemental Table B.3), showing that their mathematical model can also be applied for *P. putida* KT2440.

Durations of cell cycle phases have been shown to vary with growth conditions and nutrient availability *helmstetter1996*. Our data for *P. putida* KT2440 are in agreement with observations of Kubitschek et al. (1978) and Helmstetter et al. (1976) who detected a steady decrease of C phase length with increasing growth rate of *E. coli* B/r strains. While *E. coli* B/r reached a minimum C duration of about 42 min at growth rates $\mu > 0.7 \text{ h}^{-1}$ (generation time $\tau < 1.0 \text{ h}^{-1}$), we identified a minimum C length of about 62 min for *P. putida* KT2440 for $\mu > 0.6 \text{ h}^{-1}$ (i.e. $\tau < 1.2 \text{ h}^{-1}$) (Figure B.2). Comparing the minimum C durations of *E. coli* and *P. putida* (42 min and 62 min, respectively) with the generation time at high growth rates, multifork DNA replication needs to start already at a lower growth rate in *P. putida* than in *E. coli*.

In general, we found similar maximal replication rates r_c when comparing different unimorph Gram-negative organisms dividing symmetrically with binary fission. Michelsen et al. (2003) reported a minimum duration of the C phase of 46 min for *E. coli* K-12 and an r_c of 100 kbp/min (Myllykallio et al., 2000). Simulation of flow cytometric data obtained by Wiacek et al. (2006) in *Cupriavidus necator* resulted in a replication phase length of 83 min, thus mirroring an r_c of 102 kbp/min. For *P. putida* KT2440 we report an r_c of 100 kbp/min. Pooling these results, an average maximum replication rate of ≈ 100 kbp/min can be deduced with a very small variance of 0.6 % at standard conditions. These bacteria might share common basic properties of the replication machinery which results in similar maximum replication capacities. Noteworthy, this similarity of replication rates could not be found for asymmetric dividing organisms, e.g. *Caulobacter crescentus* (42 kbp/min (Myllykallio et al., 2000)), archaea, e.g. *Pyrococcus abyssi* (36 kbp/min (Myllykallio et al., 2000)), or *Mycoplasma capricolum* (12 kbp/min (Seto et al., 1998)) (Rocha, 2004). Nevertheless, the finding of similar maximal replication speeds among this diverse group of bacteria is intriguing.

Our results show, that the cell cycle is altered substantially under stressed conditions. The time for the replication of the chromosome was shortened and, accordingly, r_c gradually increased with the severity of the stress up to 1.9 fold. In addition to previously described alterations in the cell cycle under limiting conditions (Müller, 2007; Cooper, 1991), we found that the time before start of replication (B phase) and the time after completion of replication until division (pre-D / D phase) extended on the expense of the duration of replication itself (C phase).

The B phase is already described to adjust to different growth situations and even to vanish when conditions are optimal (Helmstetter, 1996). Therefore, it is not surprising that the B phase covers part of the surplus cell cycling time under stressful conditions at a constant μ of 0.2 h^{-1} . Regarding the cell division, the classical cell cycle model of Cooper and Helmstetter (1968) suggested that the duration of the D period is fixed. This idea might be supported by the fact that the macromolecule machinery of the divisome mechanically performs the separation of the daughter from the mother cell; a process whose interruption can be fatal for the cell (Adams et al., 2009; Huang et al., 2013).

However, our results suggest that the time between end of replication and final division (pre-D / D period) is variable and extends under stress conditions. Different explanations can be suggested for that, which go hand in hand with the existence of a pre-D period: For example, the delayed division could be a direct consequence of lower availability of resources which are re-distributed by the cell in favor of DNA replication. Cells that do not divide under limiting conditions are a common observation in batch experiments: A gap between end of replication and start of division was already suggested for limiting conditions for several bacterial species, leading to the introduction of the pre-D phase into the bacterial cell cycle model (Müller, 2007). In addition, for some archaea (Lindås et al., 2013; Hjort et al., 2001) and bacteria (Robert et al., 2014) the need of an enlarged, cell size generating period between end of replication and final division was demonstrated.

Our data show a clear relationship between acceleration of replication and general stress. We propose, that this acceleration is an actively regulated process, as seen by the higher expression of genes involved in DNA replication. This is supported by previous assumptions, that (i) replication might not proceed at maximum velocity to assure stable and correct replication and that (ii) faster replication might be achieved by a higher availability of replication processivity factors (Morigen et al., 2003; Atlung et al., 2002).

Interestingly, the expression profile also showed an upregulation of genes connected to DNA repair, many of them (*recB*, *recD*, *rwvC*) being responsible for homologous recombination. Thus, the cells might try to evade stress by a twofold strategy: To repair stress-induced errors in DNA as good as possible, while taking into account a higher frequency of recombination events, that may help a population to evolutionary adapt to challenging stress conditions more quickly.

Our study demonstrated that acceleration of DNA replication is an orchestrated cellular process between the cell cycle phases B, C, pre-D and D. Fast replication of the genetic information turned out to be of utmost priority under stress conditions. This process is balanced by extending the duration of the B and pre-D phases, which seems to be a cellular strategy to cope with stress while maintaining a constant growth rate.

B.5. Acknowledgments

This work was supported within the ERA-IB / ERA-NET scheme of the 6th EU Framework Programme (0315932B).

B.6. Supplemental material

Supplemental dataset 1

The supplemental dataset 1 contains the differentially expressed genes under decanol stress exposure in comparison to the reference, non-stressed condition. The dataset can be found on the data carrier, attached to this thesis (AppendixB_SupplementalDataset1.xlsx).

Supplemental file S1 – The chemostat

Introduced simultaneously by Novick and Szilard (1950) and Monod (1950) the chemostat is the most commonly used experimental approach for investigations of physiology in steady state cultures (Bull, 2010). The growth rate of an organism is dependent on the nutrient availability as formulated by Monod (1949):

$$\mu = \mu_{max} \frac{S}{K_S + S}$$

In 1956, Herbert published that keeping the substrate concentration at a certain level, one can vary the growth rate of an organism externally. In chemostats the dilution rate D is a function of the flow rate F and the cultivation volume V as follows:

$$D = \frac{F}{V}$$

In the bioreactor, biomass formation equals wash-out with a dilution rate D for steady state condition ($dX/dt = 0$):

$$\frac{dX}{dt} = \mu X - DX \stackrel{!}{=} 0$$

Hence, establishing steady state conditions ($dX/dt = 0$) results at equal growth rate μ and dilution rate D . It is a key property of chemostats to limit cell growth for instance by the availability of the carbon source (here: glucose) while leaving other operating parameters (such as pH, temperature etc.) constant or at saturating levels (other media components). In consequence, it is possible to evaluate specific parameter influences by fixing the growth rate and keeping all other parameters constant (Winder et al., 2011). A putative drawback of the experimental set-up was re-emphasized by Ferenci (2006): the nutrient-limited conditions could increase the selection for mutations. Therefore, it is advisable to carefully monitor the population before and after the environmentally changes for obvious differences.

Supplemental file S2 – Mathematical model

The mathematical model for calculating the distribution of single cells with distinct DNA content in a population is based on the Cooper-Helmstetter model (1968): The theoretical DNA histogram $n(G)$ results from linking an age distribution $n(a)$ of a population with constant growth parameters to a cellular DNA accumulation function.

The age distribution was implemented as a probability density function $n(a)$:

$$n(a) = 2 \cdot \ln 2 \cdot e^{(-a \cdot \ln 2)} \quad 0 \leq a \leq 1$$

$$\int_0^1 n(a) da = 1 \quad 0 \leq a \leq 1$$

Here, a is denoted as age and $n(a)$ represents the probability density function of a single cell in a population (Lindmo, 1982).

As a consequence of binary cell division, there has to be the double amount of new born cells in comparison to dividing cells. Newly divided cells are defined to be at age $a = 0$, while dividing ones possess age $a = 1$. The number of cells belonging to an age interval a_i to a_{ii} can be calculated by integrating $n(a)$ within the limits a_i and a_{ii} .

Cooper et al. (1968) assumed that the movement of the replication fork along the chromosome is constant, which determines that DNA synthesis starting at a given replication point is also constant irrespective of the cell cycle period. During the cell cycle of a single cell, the rate of DNA synthesis is described mathematically in a step function with two discontinuities: the initiation and termination of the DNA synthesis. The specific events when initiation and termination occur (a_1 and a_2 , respectively) are modeled as follows (Cooper et al., 1968):

$$a_1 = (x\tau - (C + D))/\tau$$

$$a_2 = (\tau - D)/\tau$$

Here, parameter x refers to multiples of generation time in which replication C and division D take place. Parameter τ refers to the generation time. To derive the amount of DNA (G) per cell at a specific age, the division cycle is divided into three periods, defined by the ages a_1 and a_2 at the discontinuities. The chromosome content can be calculated for each of these intervals as follows, considering $G(a = 1) = 2G(a = 0)$:

$$G(a) = k(F_1 a + F_3) + a_1 k(F_1 - F_2) + a_2 k(F_2 - F_3) \quad 0 \leq a \leq a_1$$

$$G(a) = k(F_2 a + F_3) + 2a_1 k(F_1 - F_2) + a_2 k(F_2 - F_3) \quad a_1 \leq a \leq a_2$$

$$G(a) = kF_3(a + 1) + 2a_1 k(F_1 - F_2) + 2a_2 k(F_2 - F_3) \quad a_2 \leq a \leq 1$$

Here, F refers to the number of replication forks in the interval i . k is the constant rate of DNA synthesis per replication fork, which can be derived by $k = \tau/2C$. The DNA distribution $n(G)$ is derived by the combination of the derivation of theoretical chromosome content (dG/da) and the age distribution $n(a)$.

$$n(G)/dG = n(a)/da$$

A step-by-step illustration of the calculation of the DNA distributions can be found in Figure B.4.

Accounting for variation in generation time of individual cells and error in measurements. The DNA histograms $n(G)$ derived from the simulation routine reflect an ideal population in which every cell exhibits the same growth rate. However, experimental ‘noise’ resulting at variations of generation times needs to be taken into account as following:

Biological variation was simulated by slight variation of the generation time τ (coefficient of variation CV=5 %) which was mirrored by an artificial division of the population into 30 subpopulations covering the total range of variance. The implementation was based on Skarstad et al. (1985). One resulting DNA distribution for the whole population is calculated containing all 30 simulated subpopulations.

Technical measurement variation was taken into account by assuming each DNA value in the DNA histogram to be normally distributed. The mean coefficient of variation was calculated as 5 %.

Calculation of the duration of cell cycle phases. Inputs for the simulation software implemented are the generation time τ and the experimentally derived DNA histograms ($n(G_{exp})$). The output values D' and C (both in hours, h) are identified via a least-square fit, minimizing the discrepancy between simulated $n(G)$ and measured $n(G_{exp})$ DNA histograms. Lower bounds were set to 0, while the upper bounds were set to $D = \tau$ and $C = \tau/0.45$, respectively (Cooper et al., 1968).

To evaluate our simulations we calculated the deviation s using the formula presented by Skarstad et al. (1985):

$$s = \sqrt{\frac{\sum_{i=1}^m (\sqrt{n(G_{exp})_i} - \sqrt{n(G)_i})^2}{m-1}}$$

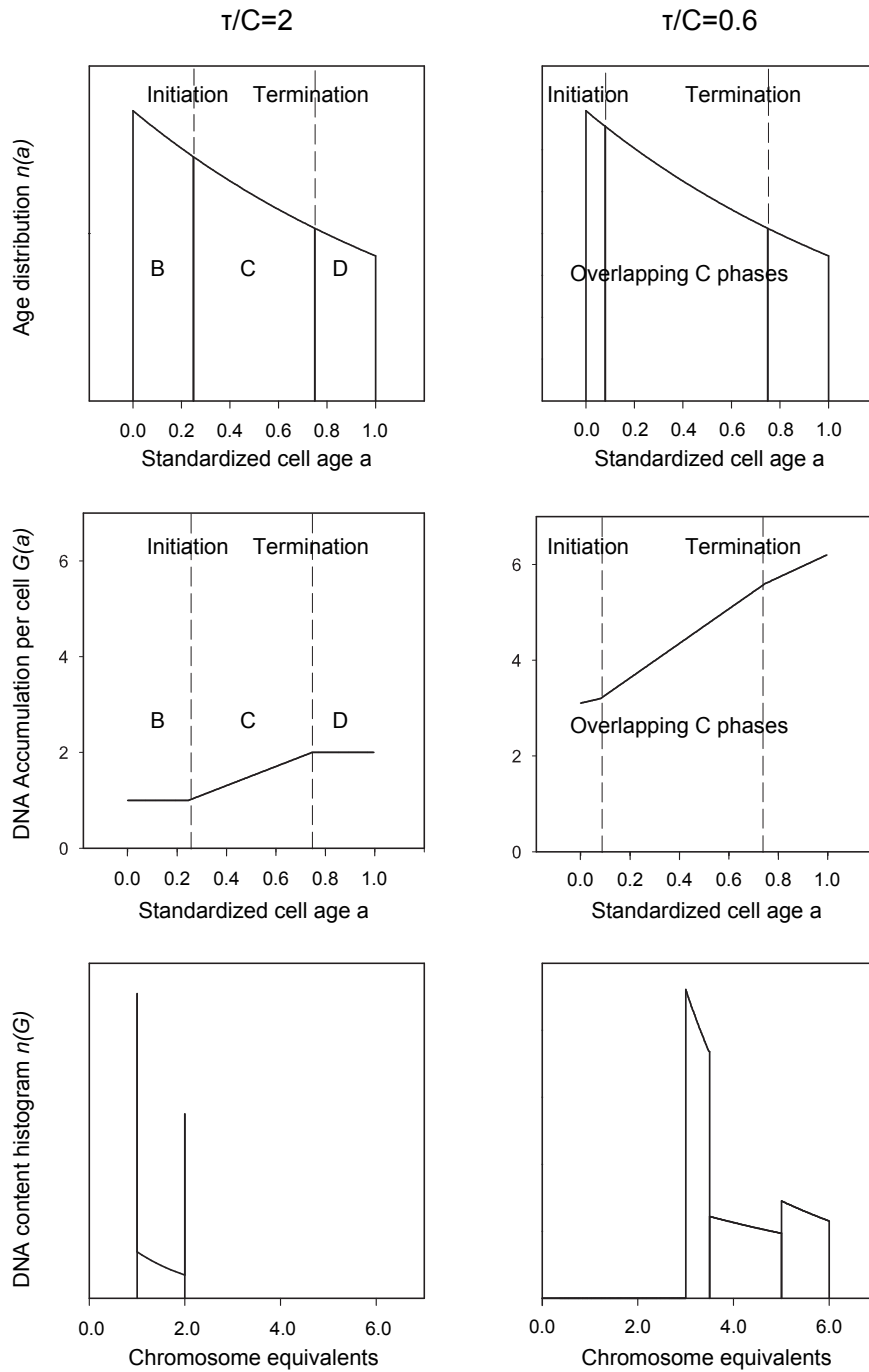


Figure B.4.: Illustration of the calculation of DNA histograms $n(G)$. DNA histograms $n(G)$ are exemplarily calculated according to Cooper and Helmstetter (1968) for a slow growing ($\tau/C = 2$) and a fast growing population ($\tau/C = 0.6$). The age distribution $n(a)$, the DNA accumulation per cell $G(a)$ and the theoretical DNA histogram $n(G)$ are illustrated in the first, second and third line, respectively. Considering the mechanism of binary fission, the number of cells that has just divided doubles those that start to divide. Depending on how fast the cells are growing, the initiation and the termination of the replication shift within the timeline of a standardized cell age a . In slowly growing cells there are phases without active replication (B and D') resulting in a constant DNA content. During the replication phase the DNA content is increasing linearly with the constant rate of replication (line 2). In the case of fast growing cells, overlapping replication cycles occur, resulting in active replication throughout the age of the cell. Replicating cells of different ages can therefore have different total replication rates according to the number of replication forks at work. The portion of cells for each DNA channel in a histogram $n(G)$ can be calculated by combining the age distribution $n(a)$ and the DNA accumulation $G(a)$ by equation $n(G)/dG = n(a)/da$ (line 3).

Supplemental Table

Table B.3.: Summary of the duration of cell cycle phases and goodness of fit of the simulation. Average values of calculated \hat{B} , \hat{C} and \hat{D}' phases (h) of 3 biological replicates and their standard deviation were calculated on the basis of the mathematical model of Cooper and Helmstetter (1968). s is the deviation of the simulated to the experimental number of cells measured by flow cytometry and presented as subpopulation distributions in DNA histograms. The formula was framed by Skarstad et al. (1985)

Stress cultivations - constant $\mu = 0.2 \text{ h}^{-1}$				
iron - 50 %	1.07 ± 0.03	1.54 ± 0.04	0.79	0.44 ± 0.26
pO_2 - 5 %	1.05 ± 0.03	1.50 ± 0.05	0.85	0.38 ± 0.24
pO_2 - 1.5 %	0.94 ± 0.05	1.44 ± 0.05	1.02	0.48 ± 0.16
decanol - 5 % (v/v)	0.80 ± 0.04	1.42 ± 0.05	1.18	0.36 ± 0.15

Supplemental Figure

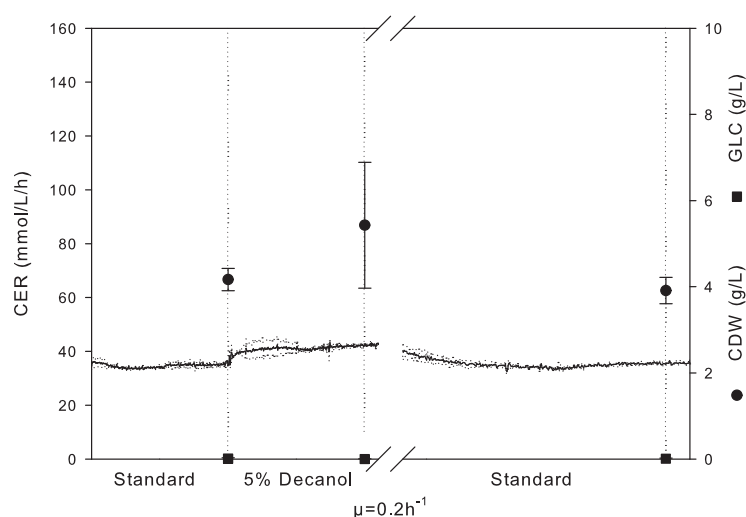


Figure B.5.: Summary of the time course of the solvent stress chemostat. The dilution rate was kept constant at $\mu = 0.2 \text{ h}^{-1}$. The solvent stress was introduced puls-wisely. After the stress condition, the culture was shifted back to reference conditions. Samples were taken at steady state conditions. Biomass densities (black circles, g L^{-1}) and residual glucose concentrations (black squares, g L^{-1}) were measured. The carbon dioxide emission rate (CER, black line) was monitored online. Error bars and lines (CER, gray dotted line) represent the standard deviation of independent biological triplicates. The relatively large error bar for CDW measurement in decanol conditions was caused by superimposing effects of the organic solvent on the gravimetric biomass detection. Note that this measurement was independently assured by cell counting.

C. Manuscript III

The implementation of novel biotechnological platform cells for industrial applications is currently the subject of intense research. Recent efforts included the use of *Pseudomonas putida* KT2440 as the functional chassis for targeted genomic manipulations aimed at reducing its extant genome. The excised functions included flagellar motility and a number of genes expected to enhance genotypic and phenotypic stability of the cells upon deletion. In this study, two multiple-deletion *P. putida* strains were evaluated as cell factories for heterologous protein production and compared to the parental bacterium in respect to several industrially-important physiological traits. Energetic parameters were quantified at different controlled growth rates in continuous cultivations and both strains had a higher adenosine triphosphate content and adenylate energy charge than the wild-type strain KT2440. Under all the conditions tested, the mutants also grew faster and had enhanced biomass yields. In addition to small scale shaken-flask cultivations, the performance of the genome-streamlined strains was evaluated in larger scale bioreactor batch cultivations taking a step towards industrial growth conditions. When the production of the green fluorescent protein was assessed in these cultures, the mutants reached a recombinant protein yield on biomass up to 40 % higher than that of *P. putida* KT2440. Taken together, the results demonstrate that these genome-streamlined derivative strains are not only robust microbial cell factories, but also a promising foundation for further biotechnological applications.

C.1. Introduction

Much of contemporary metabolic engineering approaches, both at the laboratory scale and in industrial setups, mostly rely on the use of a few bacterial hosts as working platforms (Danchin, 2012; Singh, 2014). However, the organisms that are easiest to manipulate are often not the most suitable or the most appropriate for specific large-scale and industrial applications. Several physiological and metabolic traits are desired in a robust production host (Almquist et al., 2014; Foley et al., 2010; Sauer et al., 2012). In the first place, the platform cells must be hefty and able to endure a suite of environmental and process-related stresses (Hoffmann et al., 2004). Whenever possible, the cells should also exhibit decreased (and traceable) genetic drift, physically robust envelopes, efficient and as-simple-as-possible transcription and translation controls, and predictable metabolic behavior (Foley et al., 2010). Furthermore, the concept of a suitable host for biotechnological applications is reminiscent to that of a minimal microbial cell, in which all the elements deemed unnecessary for cellular functions other than replication and self-maintenance (e.g., prophages, flagellar genes, cell-to-cell communication devices) have been eliminated. In spite of the evident

This chapter has been published as: Sarah Lieder[#], Pablo I. Nikel[#], Víctor de Lorenzo and Ralf Takors (2015) Genome reduction boosts heterologous gene expression in *Pseudomonas putida*. *Microbial Cell Factories* 14:23 [#] *Ex aequo* contribution

need for a bacterial chassis reuniting most of these desirable traits, only few hosts [typically *Escherichia coli* strains (Chen et al., 2013; Gopal et al., 2013; Jana et al., 2005; Mizoguchi et al., 2007; Ruiz et al., 2013)] are considered suitable as biocatalysts in relevant industrial endeavors, such as the production of functional recombinant proteins.

Building on the concepts outlined above, we advocate the choice of *Pseudomonas putida* strains as microbial platforms pre-endowed with metabolic and stress-endurance traits that are optimal for biotechnological needs (Nikel et al., 2014a). In particular, the non-pathogenic *P. putida* strain KT2440 shows a remarkable metabolic diversity, amenability to genetic manipulation, and stress endurance, along with the welcome GRAS (generally regarded as safe) status (Nogales et al., 2008; Poblete-Castro et al., 2012; Kim et al., 2014). Sequencing of the 6,181,863-bp long genome of *P. putida* KT2440 brought forth a significant advance in the potential applications of this bacterium (Nelson et al., 2002; Weinel et al., 2002). In an effort to enable the analysis of strain KT2440 from a systems biology perspective and to foster the development of its biotechnological applications, multiple tools for genome editing have been devised and implemented (Martínez-García et al., 2011a; Martínez-García et al., 2011b; Silva-Rocha et al., 2013). These tools have facilitated the design of a number of streamlined-genome (SG) variants derived from the wild-type strain. For instance, the construction and physiological characterization of a flagella-less variant of *P. putida* KT2440 with some attractive emergent properties, such as an elevated NADPH/NADP⁺ redox ratio, was recently reported by Martínez-García et al. (2014b). Likewise, the physiological effects of freeing the bacterium of all the viral DNA encoded in its extant chromosome (represented by not less than four prophages) was explored in several mutants (Martínez-García et al., 2014a). While such genetic manipulations conferred interesting biotechnological properties to the bacterial chassis, the industrial worth of a reduced genome *P. putida* strain has not been systematically explored hitherto. As a matter of fact, the rational engineering of cell factories tailored for optimized protein synthesis and process performance, low energy demands, and high production yield has traditionally been focused on biochemical engineering aspects (i.e., bioreactor setup and control) rather than improving the biocatalyst properly.

In this study, we have assessed the use of two heavily re-factored *P. putida* strains (one of them lacking flagella, and the other one carrying multiple mutations implemented to ensure genetic and physiological stability, see Figure C.1) as potential hosts for protein production in a bioreactor setup. The well-known green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* was selected as a model protein (Vizcaino-Caston et al., 2012), and kinetic and physiological parameters related to cell performance were analyzed in both, batch and continuous cultures. The two re-factored versions of *P. putida* KT2440 outcompeted their parental strain in every parameter tested, showing improved resistance to stress and enhanced protein production.

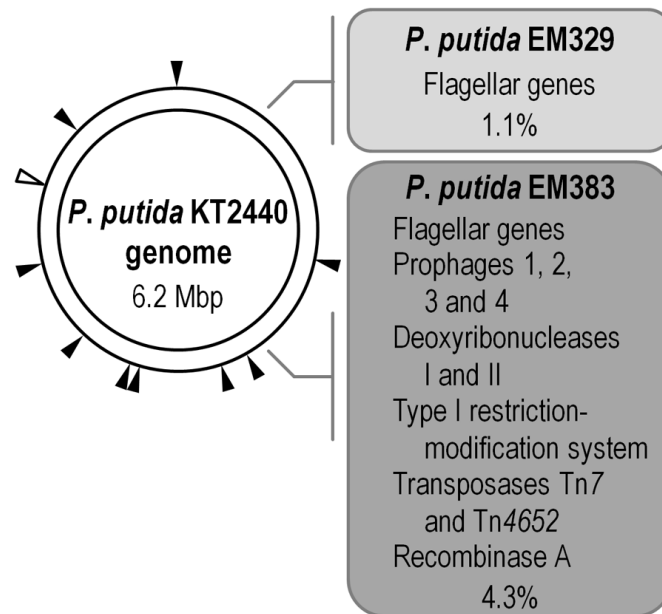


Figure C.1.: Rationale behind the design of reduced-genome derivatives of *P. putida* KT2440. Strains EM329 and EM383 were constructed using the seamless deletion system described by Martínez-García and de Lorenzo (2011a). Note that, while strain EM329 only lacks the genes encoding flagellar genes (Martínez-García et al., 2014b), the multiple deletions in strain EM383 were designed to endow the bacterium with the properties of a true microbial platform for a variety of applications. The relative physical location of the genes eliminated in the chromosome of *P. putida* KT2440 are indicated with arrowheads and the percentage of the genome deleted is shown in each case. The white arrowhead represents the chromosomal location of the flagellar genes (deleted in strain EM329), while the black arrowheads indicate the genes and gene clusters eliminated in strain EM383.

C.2. Material and Methods

Bacterial strains, culture media, and general procedures

Bacterial strains and plasmids used in this study are listed in Table C.1. *E. coli* and *Pseudomonas* strains were routinely grown at 37 °C and 30 °C, respectively, in rich LB medium (Green et al., 2012) under oxic conditions (i.e., in Erlenmeyer flasks containing medium up to one-tenth of their nominal volume with agitation at 170 r.p.m.). *E. coli* DH5 α was used for routine cloning procedures and plasmid maintenance. The physiological characterization of *P. putida* recombinants was carried out both in shaken-flask and bioreactor cultures using M12 minimal medium, which contained 2.2 g L⁻¹ (NH₄)₂SO₄, 0.4 g L⁻¹ MgSO₄ · 7H₂O, 0.04 g L⁻¹ CaCl₂ · 2H₂O, 0.02 g L⁻¹ NaCl, 2 g L⁻¹ KH₂PO₄, added with trace elements (2 mg L⁻¹ ZnSO₄ · H₂O, 1 mg L⁻¹ MnCl₂ · 4H₂O, 15 mg L⁻¹ Na₃-citrate · 2H₂O, 1 mg L⁻¹ CuSO₄ · 5H₂O, 0.02 mg L⁻¹ NiCl₂ · 6H₂O, 0.03 mg L⁻¹ Na₂MoO₄ · 2H₂O, 0.3 mg L⁻¹ H₃BO₃, 10 mg L⁻¹ FeSO₄ · 7H₂O). All cultivations were started using cells from a single colony in an LB plate, grown and harvested from exponential phase cultures in LB medium, and stored as a working cryo-culture bank at -70 °C in a 20 % (v/v) glycerol stock. Glucose or citrate were used as representative glycolytic or gluconeogenic carbon sources, respectively, throughout this study. The concentration of each carbon source in pre-cultures was 4 g L⁻¹, while in batch cultivations (both in shaken-flasks and bioreactors) it was increased up to 10 g L⁻¹. All solid media used in this work contained 15 g L⁻¹ agar, and, whenever needed, kanamycin was added at 50 mg L⁻¹ as a filter-sterilized solution for plasmid maintenance. Isopropyl- β -D-thiogalactopyranoside (IPTG) was added at 1 mM to induce the expression of genes under the control of LacI^Q/P_{trc}. Growth was estimated in a Ultrospec 3000 pro UV/Visible spectrophotometer (GE Healthcare Bio-Sciences Corp., Piscataway, NJ, USA) by measuring the optical density at 600 nm (OD₆₀₀) after diluting the culture as necessary with 9 g L⁻¹ NaCl. In bioreactor cultivations, the cell dry weight (CDW) was measured in culture aliquots as appropriate for further mass-based calculations. CDW was determined in 10 mL culture samples by transferring the broth into previously-weighed glass tubes. The suspension was centrifuged at 7,000 r.p.m. and 4 °C for 10 min and washed twice with 5 mL of cold saline. The pellet fraction was finally dried at 85 °C until constant weight (ca. 48 h). The yield of biomass on substrate ($Y_{X/S}$, in g_{CDW}g_{glucose}⁻¹) was derived from the CDW assessed in the samples and the glucose consumption rates (see below).

Bioreactor cultures

All bioreactor cultures were carried out in an in-situ sterilizable KLF 3.7-liter fermentor (Bioengineering AG, Wald, Switzerland). Exhaust gas composition (CO₂ and O₂), dissolved O₂ concentration, and pH in the liquid phase were monitored online using BCP-CO₂ and BCP-O₂ analyzers (BlueSens GmbH, Herten, Germany) and O₂ and pH probes (Mettler Toledo GmbH, Giessen, Germany). The exhaust gas measurement of CO₂ was used to calculate CO₂ emission rates. The

Table C.1.: Bacterial strains and plasmids used in this study

	Relevant characteristics ^a	Source or reference
<i>E. coli</i>		
DH5 α	Cloning host; F- λ - <i>endA1 glnX44</i> (AS) <i>thiE1</i> <i>recA1 relA1 spoT1 gyrA96</i> (Nal ^R) <i>rfbC1</i> <i>deoR nupG</i> Φ 80(<i>lacZ</i> Δ M15) Δ (<i>argF-lac</i>)U169 <i>hsdR17</i> (<i>r_K</i> - <i>m_K</i> ⁺)	Hanahan et al., 1980
<i>Pseudomonas putida</i>		
KT2440	Wild-type strain, spontaneous restriction-deficient derivative of strain mt-2 cured of the TOL plasmid pWW0	Bagdasarian et al., 1981
EM329	Flagella-less derivative of KT2440; Δ PP4329-PP4397 (flagellar operon)	Martínez-García et al., 2014a
EM383	Streamlined derivative of KT2440; Δ PP4329-PP4397 (flagellar operon) Δ PP3849-PP3920 (prophage 1) Δ PP3026-PP3066 (prophage 2) Δ PP2266-PP2297 (prophage 3) Δ PP1532-PP1586 (prophage 4) Δ Tn7 Δ <i>endA-1</i> Δ <i>endA-2</i> Δ <i>hsdRMS</i> Δ flagellum Δ Tn4652	Martínez-García et al., submitted 2014
Plasmid		
pSEVA234b	Expression vector; <i>oriV</i> (pBBR1) <i>lacI^Q</i> <i>P_{trc}</i> <i>aphA</i> , Km ^R	Silva-Rocha et al., 2013
pSEVA637b	Cloning vector carrying the green fluorescent protein gene; <i>oriV</i> (pBBR1) <i>aacC1</i> , Gm ^R	Silva-Rocha et al., 2013
pS234G	Expression vector carrying the green fluorescent protein gene under control of the inducible <i>P_{trc}</i> promoter; <i>oriV</i> (pBBR1) <i>lacI^Q</i> <i>P_{trc}</i> \rightarrow <i>gfp aphA</i> , Km ^R	This study

^a Antibiotic markers: Gm, gentamicin; Km, kanamycin.^b Plasmids belonging to the SEVA (Standard European Vector Architecture) collection.

dissolved O_2 concentration was monitored to assure non-limiting aerobic conditions. In all cultivations, the dissolved O_2 level was kept higher than $pO_2 = 70\%$ ($pO_2 = 100\%$ was defined as the dissolved O_2 level in the bioreactor under operating conditions, but without biomass in suspension). A pre-culture was prepared for each run by inoculating cells from a working cryo-culture bank (8.5 mL) in 150 mL of M12 minimal medium contained in a 1.5-liter baffled Erlenmeyer flask. Cells were cultivated as explained above until the culture reached $OD_{600} = 1.5$ and used as the inoculum as follows.

Batch cultivation

Bioreactor batch cultivations were inoculated aseptically with the mid-exponential, shaken-flask pre-culture to reach a final working volume of 1.5 liters. Previous to inoculating the bioreactor, the operating conditions were set to $30\text{ }^\circ\text{C}$, a stirrer speed of 700 r.p.m., an over-pressure in the vessel of 0.5 bar, and an aeration of 2 L min^{-1} filtered-sterilized ambient air. The pH was set and maintained at pH 7.0 by automatic addition of 25% (v/v) NH_4OH .

Continuous cultivation

In the case of glucose-limited continuous cultivations, the batch cultivation was switched into chemostat operation when glucose was depleted. The dilution rate (D) was increased stepwise from $D = 0.1\text{ h}^{-1}$ to 0.3 h^{-1} , and finally to 0.6 h^{-1} . Each D value, determined by feeding medium at a pre-defined flow rate, was maintained for 5 residence times under steady-state conditions before further increasing the growth rate. The weight gain of the bioreactor was constantly monitored, and a harvest pump was started whenever the weight gain exceeded 10 g. Additionally, D values were manually checked by weighing the mass of the harvest outflow within a time-span of 1 h before sampling.

Nucleic acid manipulation, plasmid construction, and plasmid stability assay

DNA manipulations followed well established protocols (Green et al., 2012). Plasmid pS234G carries the green fluorescent protein gene under transcriptional control of the IPTG-inducible P_{trc} promoter. This expression vector was constructed as follows. Plasmid pSEVA637 was digested with *Hind*III and *Spe*I, and the ca. 0.7-kb DNA fragment, spanning *gfp* preceded by a synthetic ribosome binding site, was ligated into pSEVA234 restricted with the same enzymes. The ligation mixture was transformed in *E. coli* DH5 α , and positive clones were identified in LB plates containing kanamycin. Plasmid DNA was recovered from single clones and checked by automated sequencing.

Plasmids were transferred into *P. putida* KT2440 and its derivatives by electroporation (Choi et al., 2006).

Plasmid segregational stability in cells grown in shaken-flask cultures was estimated as described by Nikel and de Lorenzo (2014b). Briefly, cultures were serially diluted in 10-fold steps in LB medium containing no antibiotics. The dilution level was estimated based on OD₆₀₀ measurements of the samples, and 50 µL of the final dilution was plated onto LB agar with and without kanamycin. Colony forming units (CFUs) were counted after 24 h of growth at 30 °C in biological triplicates. The segregational stability of pSEVA234 and pS234G was calculated by comparing CFUs in plates with and without kanamycin.

Analytical procedures

GFP quantification

Determination of GFP fluorescence by flow cytometry Cells sampled from shaken-flask cultures at the time points indicated in the text were immediately diluted with phosphate-buffered saline to an OD₆₀₀ of ca. 0.35 and fixed with 0.4% (v/v) formaldehyde. Flow cytometric analysis of GFP fluorescence levels was performed in a GalliosTM flow cytometer (Beckman Coulter Inc., Indianapolis, IN, USA) equipped with an argon ion laser of 15 mW at 488 nm as the excitation source. Size-related forward scatter signals gathered by the cytometer were analyzed using the CyflogicTM 1.2.1 software (CyFlo Ltd., Turku, Finland) to gate fluorescence data only from bacteria in the stream. The green fluorescence emission was detected using a 530/30-nm band pass filter set. Data for > 15,000 cells per experiment were collected, and the CyflogicTM 1.2.1 software was used to calculate the geometric mean of fluorescence per bacterial cell (x-mean) in each sample.

Determination of GFP fluorescence by spectrofluorimetry Fluorescence in samples from bioreactor cultures was determined by taking 200 µL technical triplicates of the cell suspension and the corresponding filtrates into a 96-well microtiter plate. The fluorescence was quantified at 485 nm (excitation) and 535 nm (emission) in a fluorescence microplate analyzer (Synergy 2, BioTek Instruments, Inc., Winooski, VT, USA). The yield of GFP on biomass ($Y_{\text{GFP}/X}$, in arbitrary fluorescence units (A.F.U.) g_{CDW}⁻¹) was derived from these measurements.

Kinetics of GFP accumulation in bioreactor cultures The trajectory of GFP increase was analyzed throughout the growth curve in batch cultures. To eliminate the maturation time as a possible error factor (e.g., due to varying growth rates and cultivation times), a factor, termed π_{max} , was implemented to describe the increase of GFP over time. This factor is analogous to π , the specific growth rate, which describes the increase of biomass over time during exponential growth. The corresponding equation is:

$$C_P = C_P^0 \cdot e^{\pi_{max} \cdot t}$$

where C_P is the GFP concentration (in A.F.U L⁻¹), C_P^0 is the GFP concentration at $t = 0$ h, π_{max} is the maximum specific rate of GFP formation (in h⁻¹), and t is time (in h).

Cell viability

We resorted to the propidium iodide (PI, a strong DNA intercalating agent) test, based on dye exclusion, to estimate the cell viability in samples from shaken-flask cultures. Cells having intact, polarized membranes are able to interact with and to exclude charged molecules like PI, while dead or seriously damaged bacteria become stained with the dye (Nikel and de Lorenzo, 2012). Flow cytometry analysis was performed to evaluate the percentage of PI-stained cells as a measure of cell viability. Measurements were performed in a GalliosTM flow cytometer (Beckman Coulter Inc.), using the argon ion laser at 488 nm as the excitation source. The characteristic PI fluorescence emission at 617 nm was detected using a 620/30-nm band pass filter array. PI (Life Technologies Corp., Grand Island, NY, USA) was used from a freshly-prepared stock solution at 0.5 gL⁻¹ in water and added to a final concentration of 1.5 mgL⁻¹ to the cell suspension. Cells were stained for 30 min in the dark, and measured thereafter.

Quantification of glucose and organic acids

The concentration of residual glucose and citrate in the supernatants was quantified using commercial kits according to the manufacturer's instructions (R-Biopharm AG, Darmstadt, Germany). The evolution of gluconate was also followed using a similar procedure, using a kit from Megazyme International Ireland (Bray, Ireland). In either case, control mock assays were conducted by spiking M9 minimal medium with different amounts of the carbon source under examination.

Determination of ATP/ADP ratios, ATP yields, and the adenylate energy charge

Biocatalytic reactions in the cells were stopped by promptly mixing the samples with 35% (w/v) HClO₄. A 4 mL sample was taken with a fast sampling probe directly into 1 mL of pre-cooled (-20 °C) HClO₄ solution on ice and mixed immediately. The sample was shaken at 4 °C for 15 min in an overhead rotation shaker. Afterwards, the solution was neutralized on ice by fast addition of 1 mL of 1 M K₂HPO₄ and 0.9 mL of 5 M KOH. The neutralized solution was centrifuged at 4 °C and 22,000 r.p.m. for 10 min to remove cell debris, and precipitated proteins and KClO₄. The supernatant was kept at -20 °C for batch high pressure liquid chromatography (HPLC) measurements. At each sampling time, a broth sample containing cells and a filtrated sample without cells was treated according to this procedure. Nucleotide analysis was performed by reversed-phase

ion-pair HPLC. The HPLC system (Agilent Technologies GmbH, Waldbronn, Germany) consisted of an Agilent 1200 series auto-sampler, binary pump, thermostated column compartment, and a diode array detector set at 260 and 340 nm. The nucleotides were separated and quantified on a reversed-phase C18 column combined with a security guard column (Supelcosil LC-18-T, 25 cm x 4.6 mm, 3 μ m particle size, equipped with 2 cm Supelguard LC-18-T replacement cartridges; Supelco Inc., Bellefonte, USA) at a constant flow rate of 1 ml min⁻¹. The mobile phases were [i] buffer A [0.1 M KH₂PO₄/K₂HPO₄, with 4 mM tetrabutylammonium sulfate and 0.5 % (v/v) CH₃OH, pH = 6.0] and [ii] solvent B [70 % (v/v) buffer A and 30 % (v/v) CH₃OH, pH = 7.2]. The following gradient program was implemented to separate the nucleotides in the samples: 100 % buffer A from 0 min to 3.5 min, increase to 100 % solvent B until 43.5 min, remaining at 100 % solvent B until 51 min, decrease to 100 % buffer A until 56 min, and remaining at 100 % buffer A until 66 min.

The adenylate energy charge (AEC) is a quantitative measure of the relative saturation of high-energy phospho-anhydride bonds available in the adenylate pool of the cell (Atkinson and Walton, 1967; Chapman et al., 1971), and can be expressed according to the formula:

$$\text{AEC} = ([\text{ATP}] + 0.5 \cdot [\text{ADP}]) / ([\text{ATP}] + [\text{ADP}] + [\text{AMP}])$$

The AEC values were derived from the experimental measurements of each adenine nucleotide in the samples. The amount of ATP available per unit of biomass ($Y_{\text{ATP}/X}$, in $\mu\text{mol ATP g}_{\text{CDW}}^{-1}$) was also calculated.

Calculation of maintenance demands

Maintenance demands on glucose (m_S , in $\text{g}_{\text{glucose}} \text{g}_{\text{CDW}}^{-1} \text{h}^{-1}$) were calculated by following the Pirt's equation (Pirt, 1965):

$$q_S = m_S + \mu / Y_{X/S, \text{true}}$$

where q_S is the specific rate of glucose consumption (in $\text{g}_{\text{glucose}} \text{g}_{\text{CDW}}^{-1} \text{h}^{-1}$), μ is the specific growth rate (in h^{-1}), and $Y_{X/S, \text{true}}$ is the true yield of biomass on glucose (in $\text{g}_{\text{CDW}} \text{g}_{\text{glucose}}^{-1}$).

A linear regression was used to calculate m_S values through a weighted least-squares regression. This method allows to take into account the variance of each data point individually, instead of assuming a constant variance. Weighted least-squares regression minimizes the error estimate (s) according to the following equation:

$$s = \sum_i \omega_i \cdot (y_i - \hat{y}_i)$$

where ω_i is the i -th weight, and y_i and \hat{y}_i are the measured data points and the data points derived from regression, respectively. The weights determine how much each value influences the final

parameter estimate (Fuller, 2009). Therefore, the fit is less influenced by data points of higher variance (σ_i^2) than sampling points with lower variance. The weights are calculated using the following equation:

$$\omega_i = 1/\sigma_i^2$$

Statistical analysis

The reported experiments were independently repeated at least twice (as indicated in the text), and, unless indicated otherwise, the mean value of the corresponding parameter \pm standard deviation is presented. All continuous cultivations were carried out in independent biological triplicates, and each sample was additionally taken in technical triplicates. Differences in results were evaluated via a two-tailed Student's t-test defining a P -value < 0.05 as significant.

C.3. Results and Discussion

Streamlined-genome *Pseudomonas putida* KT2440 as a chassis for heterologous protein production: design and construction of robust microbial cell factories

Recent efforts in designing adequate microbial cell factories have focused mostly in the deletion or insertion of a few genes that were deemed a priori candidates for manipulation. In this work, we evaluated the properties of SG strains derived from *P. putida* KT2440 under conditions compatible with both laboratory environments and industrial production. As part of a program of gene reduction, most of the elements supposedly unnecessary for the core reactions within the cells were sequentially eliminated. Figure C.1 and Table C.1 summarize the genomic deletions in each strain and the localization of these elements. In particular, strain EM383 carries extended deletions that would make it a strong candidate as a bacterial host for cloning and gene expression. While the fundamental phenotypic traits gained by deleting the genomic segments at stake have been recently documented (Martínez-García et al., submitted 2014), the pertinent question that prompted this study has been how do these mutants behave under the growth conditions and physiological regimes imposed by an industrial operation. And, importantly, can their emerging properties of the strains be exploited for improving heterologous protein production?

Enhanced process parameters and energy profile of streamlined-genome derivatives of *P. putida* KT2440 in continuous cultures

Biomass yield, carbon balances, and maintenance coefficients

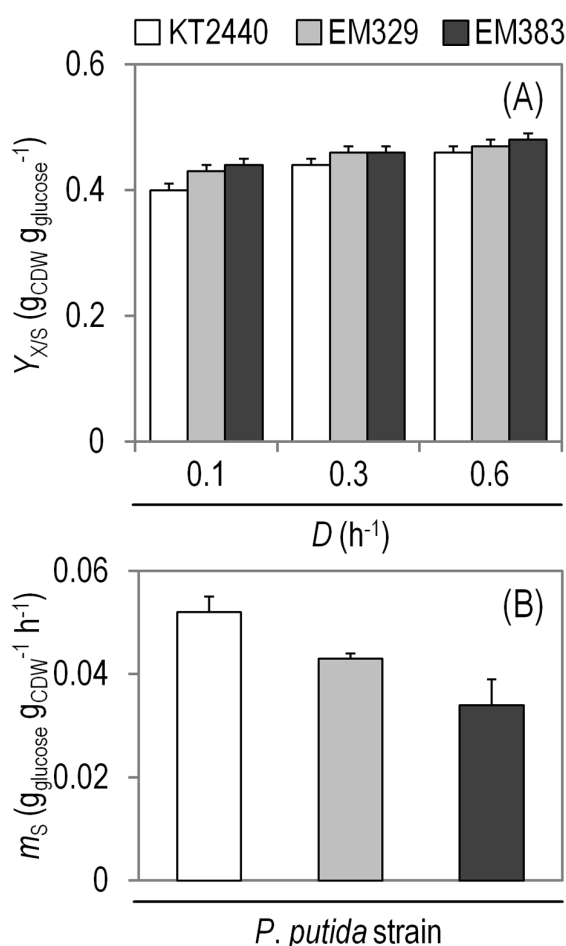


Figure C.2.: Summary of the growth parameters for the different strains under study in glucose-limited chemostat cultures. Shown are (A) the biomass yield coefficient ($Y_{X/S}$), calculated at three different dilution rates (D), and (B) the maintenance coefficient (m_S). The growth parameters were calculated based on three independent biological experiments conducted in triplicate, and the bars represent the mean value of the corresponding parameter \pm standard deviations.

processes that depend on biomass formation. The next relevant question was whether these differences in biomass yields also correlate with energy maintenance in the cognate strains.

The maintenance demand of a specific microorganism is an intrinsic characteristic of utmost im-

The starting point in the characterization of the strains under study was the setup of continuous cultivations to explore the key kinetic and process parameters of each strain at different growth rates (see section C.6 ‘Supplemental material’, Figure C.6). To this end, we started by analyzing biomass yields, a measure reflecting the efficiency of the substrate conversion into cell components. Yield coefficients were calculated in glucose-limited continuous cultivations at steady-state conditions for various D values (Figure C.2A). The mutant strains showed a higher $Y_{X/S}$ value (statistically significant, $P < 0.05$) at all growth rates when compared to the wild-type strain. The highest difference (ca. 12 %) was observed when comparing strain EM383 with wild-type KT2440 at $D = 0.1 \text{ h}^{-1}$. The differences between *P. putida* EM329 and EM383, on the contrary, were not statistically significant. The carbon emission rates (i.e., CO_2) differed significantly between the strains. Averaging over all the tested D values, strains EM329 and EM383 had 9 % and 16 % lower CO_2 evolution rates, respectively, as compared to *P. putida* KT2440. This result suggests that the carbon substrate saved by-passing the synthesis of some cellular components (e.g., flagella) can be used for macromolecular biosynthesis, accompanied by a low CO_2 evolution, an interesting trait for biopro-

portance for industrial applications and for the efficient design of production processes. As it measures the amount of carbon source (and ATP) needed to maintain minimal functions within the cell other than generation of more biomass (i.e., non-growth processes), the lower the m_S value is for a given strain and/or culture condition, the higher the carbon available to be used in catabolism (and, consequently, in biocatalysis). We explored this trait in the SG strains in the aforementioned glucose-limited continuous cultivations (Figure C.2B). Maintenance was calculated via the specific rate of glucose uptake at different D values. The linear relationship between q_S and the respective growth rate was monitored over the range of D values comprised between 0.1 and 0.6 h⁻¹. Expectedly, as D increased, so did the q_S values for each strain. By applying the Pirt's equation, an m_S of 0.052 ± 0.002 g_{glucose}g_{CDW}⁻¹h⁻¹ (corresponding to 0.29mmol_{glucose}g_{CDW}⁻¹h⁻¹) was calculated for the wild-type *P. putida* strain. Note that no by-product formation needs to be taken into account for the strains considered, as *P. putida* does not produce any excretion metabolite under these conditions (Chavarría et al., 2013; del Castillo et al., 2007). In fact, the carbon balances for all three strains showed an excellent closure (within the range $100 \pm 2\%$) just by taking into account the formation of biomass, CO₂ evolution, and the concentration of residual glucose in the culture medium (see section C.6 ‘Supplemental material’, Figure C.7).

The m_S calculated from our experimental data is in the range of the m_S values reported by van Duuren et al. (2013) for wild-type strain KT2440 in a similar chemostat setup. Vallon et al. (2013) also found m_S values in the range of those reported here when studying a *P. putida* based whole-cell biocatalysis process. The authors also pointed out that low m_S values seem to be typical for *Pseudomonas* species. For the sake of comparison with a well established bacterial host used in industrial applications, the m_S calculated for *P. putida* KT2440 in this study was ca. 28 % lower than that reported by Nanchen et al. (2006) for wild-type *E. coli* MG1655 in a similar glucose-limited continuous culture. Interestingly, the two SG counterparts of *P. putida* KT2440 had lower m_S values than their parental strain. Specifically, strains EM329 and EM383 showed a reduction in their characteristic m_S values of 17 % and 35 %, respectively, when compared to the wild-type KT2440 strain ($P < 0.01$). The corresponding $Y_{X/S,true}$ values were 0.47 g_{CDW}g_{glucose}⁻¹ for strain KT2440, and 0.49 g_{CDW}g_{glucose}⁻¹ for both EM329 and EM383. While the changes observed between the mutants and the wild-type strain were statistically significant, the difference when comparing the two SG variants was not.

In general, and according to the data available in the literature, maintenance coefficients of Gram-negative organisms grown in a defined glucose-containing medium vary from ca. 0.05 to 0.5 g_{glucose}g_{CDW}⁻¹h⁻¹ (Atkinson et al., 1967; Kooijman et al., 1992; Russell, 2007; Schulze et al., 1964). From this point of view, the calculated maintenance demands of *P. putida* and its derivatives lie within the lower end of the cited range of known m_S values. The emerging picture is that the deletion of cellular components and structures that spend energy (e.g., flagella assembly and motility) resulted in a reduction in maintenance demands in *P. putida*, therefore making the SG strains appealing production hosts. The m_S values can also be transformed into ATP demands to directly

visualize energy expenditures related to maintenance by taking into account some stoichiometric considerations. The Entner-Doudoroff pathway in *P. putida* yields 1 mole of ATP and 1 mole of NADH per mole of glucose consumed. Additionally, the tricarboxylic acid cycle forms 4 NADH and 1 FADH per each acetyl-coenzyme A, which, for the sake of simplicity in the calculations, can be lumped into 5 NADH. In consequence, 1 glucose molecule yields 1 ATP and ca. 11 NADH. Assuming a P/O ratio of 1.75, 21 ATP per glucose are formed via oxidative phosphorylation. Under these assumptions, the m_{ATP} values (in $\text{mol}_{\text{ATP}}\text{g}_{\text{CDW}}^{-1}\text{h}^{-1}$) were 1.09 ± 0.06 for *P. putida* KT2440, and 0.91 ± 0.02 and 0.71 ± 0.05 for strains EM329 and EM383, respectively, thereby mirroring the trend observed in the m_S values among the strains. According to these figures, EM329 and EM383 had a reduction of 17% and 35%, respectively, in the ATP needed for non-growth processes as compared to the parental strain. In all, these figures are likely to correlate with the reduced energy requirements due to the lack of flagella. This trait, in turn, encompasses two aspects: [i] reduced energy needs to synthesize and assemble flagellar proteins, and [ii] low energy requirements associated with flagellar operation and motility.

Energy status

During industrial production conditions, bacterial cells are constantly challenged with increased energy demands. The energetic capacity of the cells can be estimated via several physiological parameters, such as [i] the ATP/ADP ratio, [ii] the amount of ATP and the amount of total phosphorylated forms of adenine available per unit of biomass ($Y_{\text{ATP}/\text{X}}$ and $Y_{\text{AXP}/\text{X}}$, respectively), and [iii] the AEC. The AEC usually gives a deeper insight into the energy state of the cells than the ATP/ADP ratio does, because it considers the relative contribution of all three phosphorylated forms of adenine. The energy capacity of the strains under study was addressed in glucose-limited continuous cultivations at different D values (Figure C.3). At all the D values tested, strain EM383 consistently had a statistically significant increase in the ATP content and a higher AEC compared to both EM329 and KT2440 ($P < 0.01$) (Figure C.3A and C). The total amount of the three possible phosphorylated forms of adenine was also high in the mutants, and particularly in strain EM383 at $D = 0.6 \text{ h}^{-1}$ (Figure C.3B). The same phenomenon holds true when strains EM329 and KT2440 were compared side-by-side; the mutant having higher $Y_{\text{ATP}/\text{X}}$ and AEC values than the wild-type strain ($P < 0.01$). Notably, under fast growth conditions, the difference in the ATP availability between strain EM383 with respect to both EM329 and KT2440 was more than doubled (Figure C.3A). The general trend, evidenced in either reduced genome strain, was to have a higher $Y_{\text{ATP}/\text{X}}$ and an increased AEC at mid-range growth rates ($D = 0.3 \text{ h}^{-1}$) than at low ($D = 0.1 \text{ h}^{-1}$) or high ($D = 0.6 \text{ h}^{-1}$) growth rates. At the highest D, the AEC value dropped in all the strains, yet *P. putida* EM383 managed to keep a higher level of intracellular ATP even under these fast growth conditions, in clear contrast to the other two strains. These results are

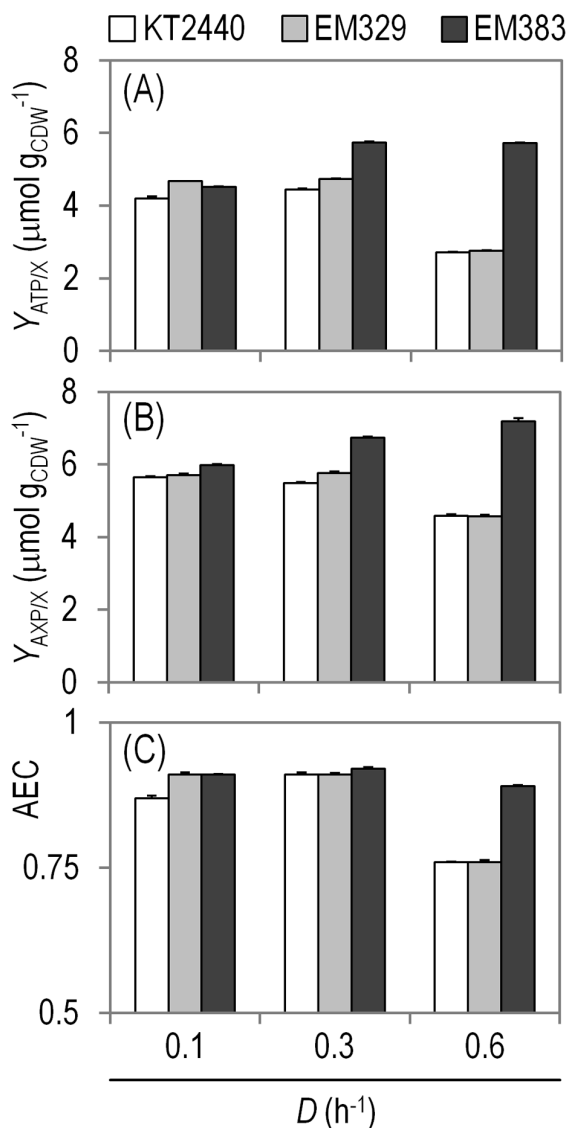


Figure C.3.: Characterization of energy parameters for the different strains under study in glucose-limited chemostat cultures. Shown are (A) the yield of ATP on biomass ($Y_{ATP/X}$), (B) the yield of total nucleosides phosphates on biomass ($Y_{AXP/X}$), and (C) the adenylate energy charge (AEC) of the cells at three different dilution rates (D). The availability of phosphorylated adenine forms inside the cell and the AEC calculations are based on three independent biological experiments conducted in triplicate, and the bars represent the mean value of the corresponding parameter \pm standard deviations.

fully consistent with the decreased maintenances in the SG strains explained above, both at the substrate and ATP demands.

Taken together, the results obtained in the glucose-limited continuous cultivations above suggested that both *P. putida* EM329 and EM383 have a number of physiological advantages over the wild-type KT2440 strain that could be potentially exploited for industrial purposes - such as expressing foraneous DNA. The systematic evaluation of these physiological traits on the background of heterologous protein production is explained in the next sections by adopting a model system which mimics industrial conditions.

Evaluation of streamlined-genome strains EM329 and EM383 as hosts for heterologous protein synthesis in batch cultures

Shaken-flask cultivation was selected as the first step in the characterization of the strains as microbial cell factories. Growth and physiological parameters as well as recombinant protein production were evaluated for each strain as explained below.

Growth parameters and kinetics of GFP accumulation

GFP was selected as the model protein to study heterologous protein synthesis in the different strains used in this study. A standardized version of *gfp*, derived from plasmid pSEVA637 (Silva-Rocha et al., 2013), was cloned into a vector in which the gene transcription is under control of an IPTG-inducible expression system (i.e., a $\text{LacI}^Q/\text{P}_{\text{trc}}$ element). The resulting plasmid, termed pS234G (Table C.1, Figure C.4A),

Table C.2.: Growth and protein synthesis parameters in shaken-flask cultures of different recombinant *P. putida* strains^a.

Strain	Plasmid ^b	Growth parameters			Protein synthesis	
		μ_{\max}^c (h ⁻¹)	CDW ^d (g L ⁻¹)	π_{\max}^c (h ⁻¹)	$Y_{\text{GFP}/X}^c$ (A.F.U. g _{CDW} ⁻¹)	
KT2440	None	0.38 ± 0.01	2.6 ± 0.9	-	-	-
	pSEVA234	0.35 ± 0.02	2.1 ± 0.3	-	-	-
	pS234G	0.28 ± 0.03	1.7 ± 0.5	0.32 ± 0.06		2,125 ± 182
EM329	None	0.47 ± 0.02	2.9 ± 0.1	-	-	-
	pSEVA234	0.45 ± 0.01	2.9 ± 0.4	-	-	-
	pS234G	0.42 ± 0.04	2.7 ± 0.3	0.41 ± 0.02		2,613 ± 107
EM383	None	0.53 ± 0.01	3.4 ± 0.2	-	-	-
	pSEVA234	0.48 ± 0.02	3.1 ± 0.5	-	-	-
	pS234G	0.46 ± 0.03	2.9 ± 0.4	0.45 ± 0.01		3,047 ± 115

^aCells were grown batchwise in M12 minimal medium containing 10 g L⁻¹ glucose as the sole carbon source and 1 mM IPTG was added in the cultures of the recombinant strains as indicated in Materials and methods. Results represent the mean value of the corresponding parameter ± standard deviation of triplicate measurements from at least two independent biological replicates.

^b Plasmid pS234G, a derivative of vector pSEVA234, carries *gfp* under control of an inducible *LacI^Q/P_{trc}* element.

^c Kinetic parameters were determined during exponential growth. μ_{\max} , maximum specific growth rate; π_{\max} , maximum specific rate of GFP formation; $Y_{\text{GFP}/X}$, yield of GFP on biomass; A.F.U., arbitrary fluorescence units; -, not applicable.

^d Final biomass concentration at 24 h. CDW, cell dry weight.

was introduced in *P. putida* KT2440 and its SG derivatives, and their behavior in shaken-flask cultures was evaluated. The impact of introducing plasmid pS234G in these strains depended on the bacterial host, as both mutants had a lower reduction in their μ_{\max} values than the wild-type did (Table C.2). In strain KT2440, introduction of the *gfp*-expressing plasmid lowered μ_{\max} in ca. 26 % when compared to the plasmid-less counterpart. In the SG derivatives, this reduction never surpassed half that value (ca. 12 %), demonstrating that the metabolic burden caused by plasmid maintenance and heterologous protein production had a low impact in strains EM329 and EM383. Both strains attained not only higher cell densities at the end of the 24-h cultivation period than KT2440, but they also grew faster irrespective of the plasmid they were transformed with. For instance, *P. putida* EM383/pS234G had an 1.6-fold increase in μ_{\max} with respect to KT2440/pS234G, and it also reached an 1.7-fold higher final CDW concentration.

Another evident difference was that GFP had a better induction profile in strains EM329 and EM383 than in wild-type KT2440 (Figure C.4). In fact, the difference between the induced versus the non-induced state in the mutants was twice as much as that observed in the parental *P. putida*

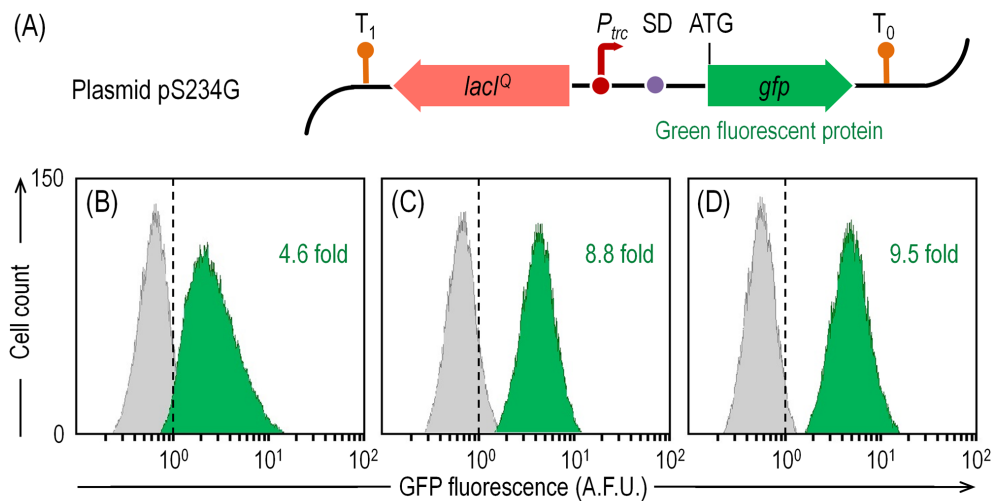


Figure C.4.: Flow cytometry analysis of the green fluorescent protein accumulation in the strains under study. (A) Schematic representation of plasmid pS234G, carrying gfp under the transcriptional control of the IPTG-inducible P_{trc} promoter. The activity of P_{trc} is controlled by the transcriptional regulator $LacI^Q$. The transcriptional terminators included in the plasmid backbone are depicted as T_0 and T_1 . The elements in this outline are not drawn to scale. *P. putida* KT2440 (B), EM329 (C), and EM383 (D) carrying pS234G were grown on M12 minimal medium containing glucose and harvested in mid-exponential phase. Gray and green peaks represent non-induced and induced cells, respectively. The vertical dashed line indicates the background fluorescence of the corresponding strain carrying the empty pSEVA234 plasmid, used as a negative control. The results shown are from a representative experiment, and the fold change in fluorescence upon induction is indicated in each case. A.F.U., arbitrary fluorescence units.

strain (Figure C.4B). The compactness of the Gaussian curves in flow cytometry experiments of both EM329 (Figure C.4C) and EM383 (Figure C.4D) also reflects a more homogenous induction of individual cells than in the wild-type strain, for which the curve in the cell counts versus GFP fluorescence plot was wider. When the trajectory of GFP formation was followed in batch cultures along the time, relevant differences were also observed (Table C.2). As previously noted, recombinant protein production is known to be proportional to growth the substrate is not limiting the growth rate. Accordingly, the maximum specific rates of GFP formation more or less paralleled μ_{max} values in each strain, with the expected result of fast GFP accumulation in the SG strains (e.g., in *P. putida* EM383, μ_{max} was 1.4-fold higher than in strain KT2440). In order to eliminate the maturation time as a possible error factor due to varying lag phases, μ_{max} and cultivation times, the cell density of the culture was correlated to the GFP fluorescence emitted. A linear regression during the exponential growth phase resulted in a correlation factor of GFP fluorescence per unit of CDW, which allowed calculating the yield of recombinant protein ($Y_{GFP/X}$). When biomass formation was also taken into account to calculate the corresponding $Y_{GFP/X}$ values, EM329 and EM383 also outcompeted *P. putida* KT2440 in 20 % and 39 %, respectively.

Enhanced cell viability of the streamlined-genome strains expressing gfp

The slight decrease in μ_{max} and in the final cell density of the recombinant strains expressing gfp

(Table C.2) suggested that the metabolic burden imposed by protein accumulation could affect final yields and the overall process performance. We asked the question of whether cell viability could be affected as well (Díaz Ricci et al., 2000), and we resorted to the PI exclusion test to explore this possibility (see section C.6 ‘Supplemental material’, Figure C.8). While *P. putida* KT2440 showed a decrease in cell viability in the presence of pS234G as compared to the same strain with an empty plasmid, neither EM329 nor EM383 showed differences in the PI staining profile. Moreover, the percentage of PI-stained cells was lower for both SG strains than for the parental host, irrespective of the plasmid they carry. Among the strains tested, *P. putida* EM383 showed the highest cell viability. Notably, when the strains bearing plasmids were compared with their plasmid-free counterparts, no decrease in cell viability was observed in strains EM329 and EM383 (data not shown). When the same comparison was established for KT2440, a significant increase (ca. 25 %) of the PI-positive population was detected in the strains carrying plasmid DNA as compared to the plasmid-free host, a figure in agreement with the results of Table C.2. These results suggest that the SG *P. putida* strains have not only a high ability of carrying and replicating heterologous plasmid DNA (see below), but also that they tolerate the metabolic burden commonly associated with plasmid replication better than wild-type *P. putida* KT2440.

Plasmid stability

All the recombinant cells were able to maintain the recombinant plasmid after 24 h of cultivation, with no significant differences among the three strains. However, when the percentage of plasmid-bearing cells was estimated after 48 h of cultivation, a significant difference in the segregational stability of pS234G could be observed. While *P. putida* KT2440 and EM329 cells retained the plasmid up to $81\% \pm 1\%$ and $85\% \pm 4\%$ of the total bacterial population, strain EM383 had a percentage of recombinants that reached $100\% \pm 2\%$ ($P < 0.05$, when compared to the other two strains). In other words, strain EM383 did not show any significant plasmid loss after prolonged cultivation, reflecting a higher stability of extra-chromosomal DNA. This phenomenon is consistent with the absence of some recombinogenic features in this strain (e.g., the Tn7 and Tn4652 transposases) that are known to bring forth genetic instability (Hörak et al., 1998; Schneider et al., 2004). Deletion of these elements results in significant genome and plasmid stabilization, which in turns is beneficial in industrial processes with long fermentation runs (Díaz Ricci et al., 2000; Soriano et al., 1999).

Kinetics of GFP formation in bioreactor batch cultures: influence of controlled aeration and carbon source on growth and profile of protein synthesis

Judging by the process parameters measured in shaken-flask cultures, the viability profile of the recombinants under these conditions, and the genetic stability of the cells, both SG strains seem

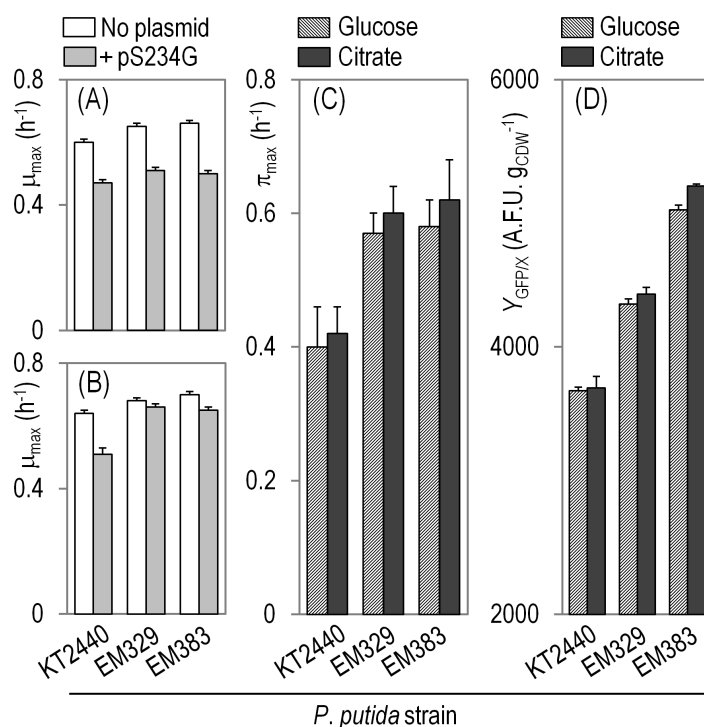


Figure C.5.: Characterization of growth parameters and protein production kinetics for the different strains under study in batch bioreactor cultures. Shown are the specific growth rate (μ_{\max}) for cells grown on (A) glucose and (B) citrate, as well as the effect of plasmid maintenance and heterologous protein production under these growth conditions. The accumulation of the green fluorescent protein (GFP) in cultures of the strains carrying pS234G was assessed during exponential growth on M12 minimal medium containing either glucose or citrate through (C) the maximum specific rate of GFP formation (μ_{\max}) and (D) the yield of GFP on biomass ($Y_{GFP/X}$). The growth parameters and protein production kinetics were calculated based on three independent biological experiments conducted in triplicate, and the bars represent the mean value of the corresponding parameter \pm standard deviations.

to be preferable over strain KT2440 as bacterial hosts for protein synthesis. We decided to further evaluate their capabilities as microbial cell factories in the well-controlled environment of a bioreactor to exploit their biotechnological potential under conditions compatible with industrial production. A detailed physiological characterization was carried out in a 3.7-liter scale bioreactor with a working volume of 1.5 liter. Growth parameters and recombinant protein production capacities were calculated for the SG strain and thoroughly compared to the wild-type counterpart.

Growth parameters

The derivative SG strains reached statistically significant higher μ_{\max} values than the wild-type KT2440 strain in all the cultivations performed (Figure C.5). When grown on glucose as the sole carbon source, EM329 showed a 7% and EM383 a 10% increase in μ_{\max} (Figure C.5A, $P < 0.05$). When using citrate as the carbon source, EM329 showed a 4% and EM383 a 11% faster growth

(Figure C.5B, $P < 0.05$). When comparing the two SG strains, mutant EM383 also reached a statistically significant higher μ_{\max} compared to strain EM329. Besides, both EM329 and EM383 attained higher final CDW concentrations when grown on glucose as the sole carbon source (9 % and 13 %, respectively when compared to *P. putida* KT2440; $P < 0.05$) (see section C.6 ‘Supplemental material’, Figure C.9), mirroring the results already observed in shaken-flask cultures (Table C.2). This difference was not observed on citrate, as all the strains reached a similar final biomass density (see section C.6 ‘Supplemental material’, Figure C.10). In all, these results show the importance of adequate aeration and mixing within the bioreactor. In the first place, all the strains attained higher μ_{\max} values and final cell densities in bioreactor cultivations as compared to the same traits in shaken-flask cultures. On the other hand, as both *P. putida* EM329 and EM383 are devoid of the flagellar machinery that would enable the cells to explore different microenvironments within the bioreactor, they tend to sediment and, if not properly stirred, the cells will likely become limited in O_2 , as previously hinted by Martínez-García et al. (2014a). The same stirring speed and air bubbling applied to the bioreactor to grow *P. putida* KT2440 enabled a much better growth profile of the SG strains.

All the strains were transformed with the expression plasmid pS234G, carrying *gfp*, to investigate recombinant protein production capacity. As a further control, the wild-type strain was also transformed with the empty vector pSEVA234. Interestingly, the introduction of the empty vector in KT2440 did not result in a significant decrease in growth (1.5 % in average), as it was also quantified in shaken-flask cultures (Table C.2). On the basis of these results, the influence of the control vector on the physiology of the cells was deemed negligible. Expression of *gfp* from plasmid pS234G, on the contrary, caused an average 6 % decrease in the μ_{\max} value for the wild-type strain. On the other hand, expression of *gfp* in the SG strains did not lead to a significant decrease of μ_{\max} . The general trend of an increase in μ_{\max} for the derivative strains previously observed in all the growth conditions analyzed could also be observed under recombinant protein expression conditions, particularly when using citrate as the sole carbon source. In fact, when growing on citrate, *P. putida* EM329/pS234G and EM383/pS234G reached significantly higher growth rates (ca. 32 % for both strains) than *P. putida* KT2440/pS234G (Figure C.5A and Figure C.5B, $P < 0.05$). No significant differences, however, were observed within the two derivative strains, as they grew very similarly and attained very comparable final cell densities.

Recombinant protein expression

During exponential growth of the cells, the trajectory of fluorescence increase due to GFP accumulation was found to be exponential as well (see section C.6 ‘Supplemental material’, Figure C.9 and Figure C.10). Under these production conditions, the μ_{\max} values were higher in the reduced genome strains as compared to the wild-type, in an almost carbon source-independent fashion

(Figure C.5C). The highest differences were detected using citrate as the carbon source; under these conditions, *P. putida* EM329/pS234G and EM383/pS234G showed an increase of 43 % and 48 % in μ_{\max} , respectively, when compared to the same parameter in *P. putida* KT2440/pS234G ($P < 0.05$). Both derivative strains also had a significantly higher $Y_{\text{GFP}/X}$ compared to the wild-type strain (Figure C.5D, $P < 0.05$), and this trend was again more or less independent of the carbon source used. For instance, when growing the cells on glucose, EM329 reached 18 % higher yield, whereas EM383 was capable of attaining a 37 % higher yield than strain KT2440. On citrate, the differences between the $Y_{\text{GFP}/X}$ values for strains EM329 and EM383 were 20 % and 41 %, respectively, as compared to wild-type KT2440. In the exponential growth phase, the volumetric productivity of GFP was estimated to be $3,470 \pm 9$ A.F.U. $\text{L}^{-1}\text{h}^{-1}$ for strain EM383/pS234G when growing on citrate, the highest among the strains and growth conditions tested in this study.

Organic acids formation

One important aspect of industrial fermentations is the spillage of by-products that divert carbon (and, most often, also cofactors such as ATP or NADPH) needed for the synthesis of the desired product (Silva et al., 2012). As mentioned in above, *P. putida* does not secrete metabolites at a high concentration, as it is the case, for example, of acetate in *E. coli* fermentations (Wong et al., 2008). However, when glucose is used as the carbon source, part of the substrate is usually oxidized by *P. putida* to gluconate in the cell periplasm by the activity of a glucose dehydrogenase (del Castillo et al., 2007). Gluconate can leak out of the cell into the culture medium and re-used as substrate as growth proceeds. When the accumulation of gluconate in the culture medium was evaluated in the bioreactor cultures, a sharp peak for *P. putida* KT2440 was observed around 5-6 h of cultivation, reaching 18.5 ± 3.1 mM (i.e., ca. 3.5 g L^{-1}). In contrast, both SG derivatives produced less gluconate during the growth phase, its concentration reaching 10.2 ± 1.4 and 9.3 ± 1.5 mM, respectively. These figures are comparable to those obtained when gluconate formation was evaluated in cultures of the strains carrying pS234G (data not shown). The kinetics of gluconate accumulation was very similar among all the strains, and this metabolite altogether disappeared from the culture supernatants after ca. 8 h as it was likely used by the cells as substrate. As a consequence of this significant reduction in the oxidation of glucose in the mutants, it is likely that more carbon is readily available for catabolism, in agreement with the high $Y_{X/S}$ values and CDW concentrations observed in the cultures of both *P. putida* EM329 and EM383.

C.4. Conclusion

Determinants of successful recombinant protein production, such as the rate and duration of production and quality or stability of the product, strongly depend on the physiology of the producer cell (de Marco, 2013). These traits can be manipulated by metabolic engineering of the host cell, by genetic engineering of the expression vector, and also by means of process engineering (Rosano

et al., 2014; Waegeman et al., 2011). However, the vast majority of metabolic engineering efforts so far have dealt with the manipulation of genetic parts implanted in a bacterial host, optimization of the process parameters, and less so with the microbial chassis itself.

At the onset of the many genome projects starting in the mid-1980s, the prevailing idea we had was that the functions required to sustain life were properly and unequivocally identified. It was therefore possible to establish a list of the minimum number of functions that would be necessary, if perhaps not sufficient, to account for the properties of living systems. Cells used to produce recombinant products are expected to accommodate artificial constructs and to behave in the predicted manner, producing the right products, with the right yield, at the right time. The emerging picture consistently shows how far we are of such ideal scenario. One of the reasons for this behavior is our current lack of knowledge about the functionality (and essentiality) of a number of genes in a wide variety of environmental conditions. In the case of the environmental bacterium *P. putida*, the set of genes strictly needed for survival in soil is most likely not the gene complement appropriate for the efficient production of heterologous proteins in an industrial setup. In particular, the deletion of the flagellar operon clearly resulted in a physiological advantage in our experimental setup, in which motility is not a required feature - and it is even a detrimental one. Bacteria had evolved fine-tuned transcriptional control mechanisms to ensure the temporal production of subsets of flagellar proteins needed for the proper flagellar biosynthesis (Chevance et al., 2008; Kazmierczak et al., 2013). The elimination of flagella in *P. putida* KT2440 determines a surplus of ATP and NADPH (Martínez-García et al., 2014b) that can be potentially funneled into a heterologous pathway. On top of the absence of the flagellar machinery, the elimination of the proviral load has been demonstrated to enhance the stress tolerance of *P. putida* KT2440 (Martínez-García et al., 2014a). In this work, the additive nature of these deletions has been exposed by exploiting the resulting bacterial chassis in a setup compatible with the industrial production of heterologous proteins.

The most prominent program of rational ‘genomic surgery’ so far has been carried out by Blattner and collaborators in the wild-type *E. coli* strain MG1655, resulting in a series of MDS derivatives (MDS standing for multiple deletion strain) that acquired advantages (mostly in terms of genetic stability) for hosting and expressing heterologous genes (Csörgo et al., 2012; Pósfai et al., 2006; Sharma et al., 2007; Umenhoffer et al., 2010). However, while significant reductions of the *E. coli* MG1655 genome size have been achieved thus far (Mizoguchi et al., 2007), these strains unavoidably retain the genomic and biochemical frame of a typical enterobacterium. This is a significant issue for expression of recombinant genes or pathways that cause stress or demand a high ATP and/or NAD(P)H availability to achieve full functionality (Na et al., 2010; Nicolaou et al., 2010), as it is the case with the streamlined *P. putida* variants examined in this article. Although the side-by-side comparison of streamlined *E. coli* and streamlined *P. putida* as microbial cell factories is beyond the scope of this work, the results presented above showed without a doubt that the two SG derivatives of *P. putida* KT2440 outcompeted the parental strain in every biotechnologically-

relevant parameter assessed among all the culture conditions tested, particularly in a bioreactor setup.

As shown above, *P. putida* EM329 and EM383 are not only sound microbial cell factories on their own, but they also provide a solid foundation for further targeted manipulations of their genomes. These forthcoming operations will not only result in enhanced bacterial chassis tailored for industrial protein synthesis, but they will also shed light on the relevant question about what is the minimal gene set needed to maintain cell functioning, fitness, and robustness. Moreover, the combination of these genomic surgery strategies along with the optimization of industrial cultivation parameters (e.g., by analyzing protein production in fed-batch cultures) will certainly result in significant improvements of the overall process performance in a variety of biotechnological applications.

C.5. Acknowledgements

The authors are indebted to E. Martínez-García (Madrid) and M. Siemann-Herzberg (Stuttgart) for sharing materials and for enlightening discussions. This study was supported by the ST-FLOW, *Pseudomonas* 2.0 (0315932B) and ARISYS Contracts of the EU, the BIO Program of the Spanish Ministry of Economy and Competitiveness, and the PROMT Project of the CAM. PIN is a researcher from the Consejo Nacional de Investigaciones Científicas y Técnicas (Argentina) and holds a Marie Curie Actions Program grant from the EU (ALLEGRO, UE-FP7-PEOPLE-2011-IIF-300508). Authors declare no conflicts of interest.

C.6. Supplemental material

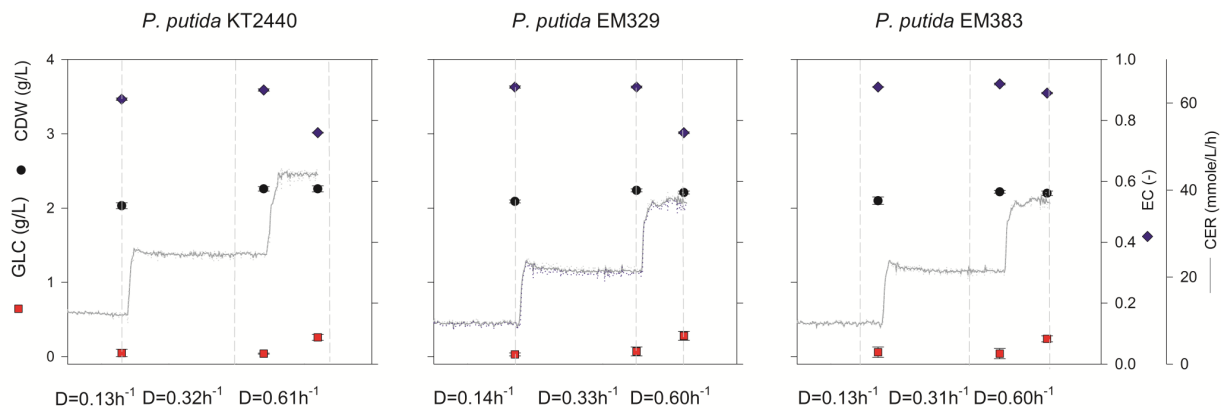


Figure C.6.: Physiological characterization of (A) *P. putida* KT2440, (B) *P. putida* EM329, and (C) *P. putida* EM383 in glucose-limited chemostat cultures at different dilution rates (D). Each cultivation was performed in biological triplicates. D was increased step-wise from D = 0.1 to 0.3 and 0.6 h⁻¹ after five residence times at each D value when a steady state was achieved. Steady states were monitored by the stable carbon emission rate (CER, black line) and stable optical density measurements (data not shown). Cell dry weight (CDW, black dots), residual glucose concentration (GLC, red squares), and the adenylate energy charge (EC, blue diamonds) were measured at steady state conditions after 5 residence times of one specific dilution rate. Error bars represent standard deviations of the biological triplicates.

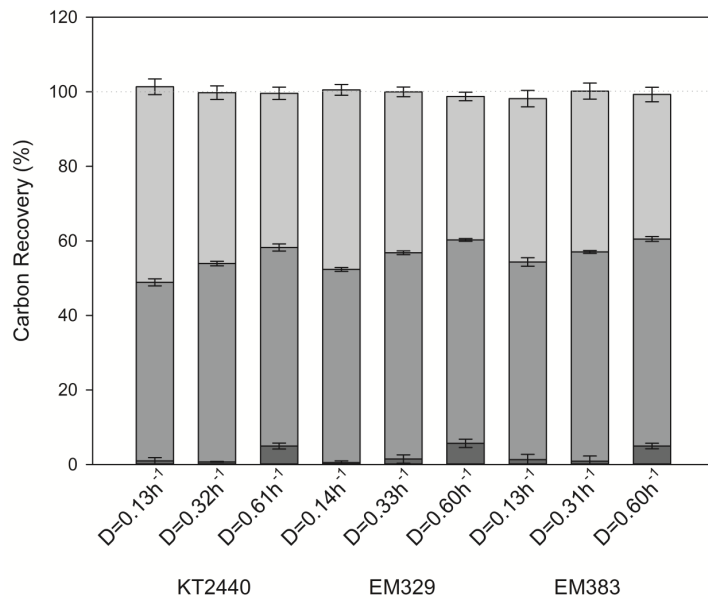


Figure C.7.: Carbon balance of glucose-limited chemostat cultures of *P. putida* KT2440, *P. putida* EM329, and *P. putida* EM383. The carbon provided by glucose served as the 100 % carbon input into the cultivation. Carbon recovery (%) was calculated considering residual glucose concentrations (dark grey), cell dry weight concentrations (grey), and CO₂ emission (light grey). Error bars represent standard deviations of the biological triplicates.

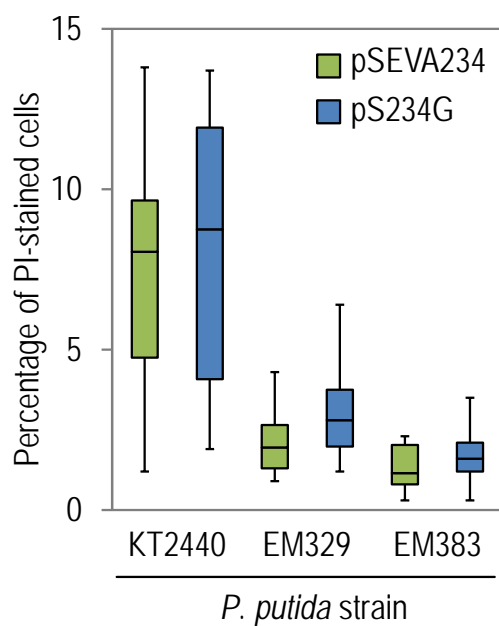


Figure C.8.: Propidium iodide (PI) exclusion was used to estimate cell viability in *P. putida* KT2440, *P. putida* EM329, and *P. putida* EM383 with the empty and the recombinant plasmid. Appropriate dilutions of cell suspensions grown on M12 minimal medium with 10 g L^{-1} glucose were stained with PI and the percentage of PI-positive cells was determined by flow cytometry as detailed in the Material and Methods section. Box plots represent the median value and the 1st and 3rd quartiles of the geometric mean values of quadruplicate determinations from three independent cultures, and the asterisks identify significant differences at the $P < 0.05$ level as assessed with the Mann-Whitney U test.

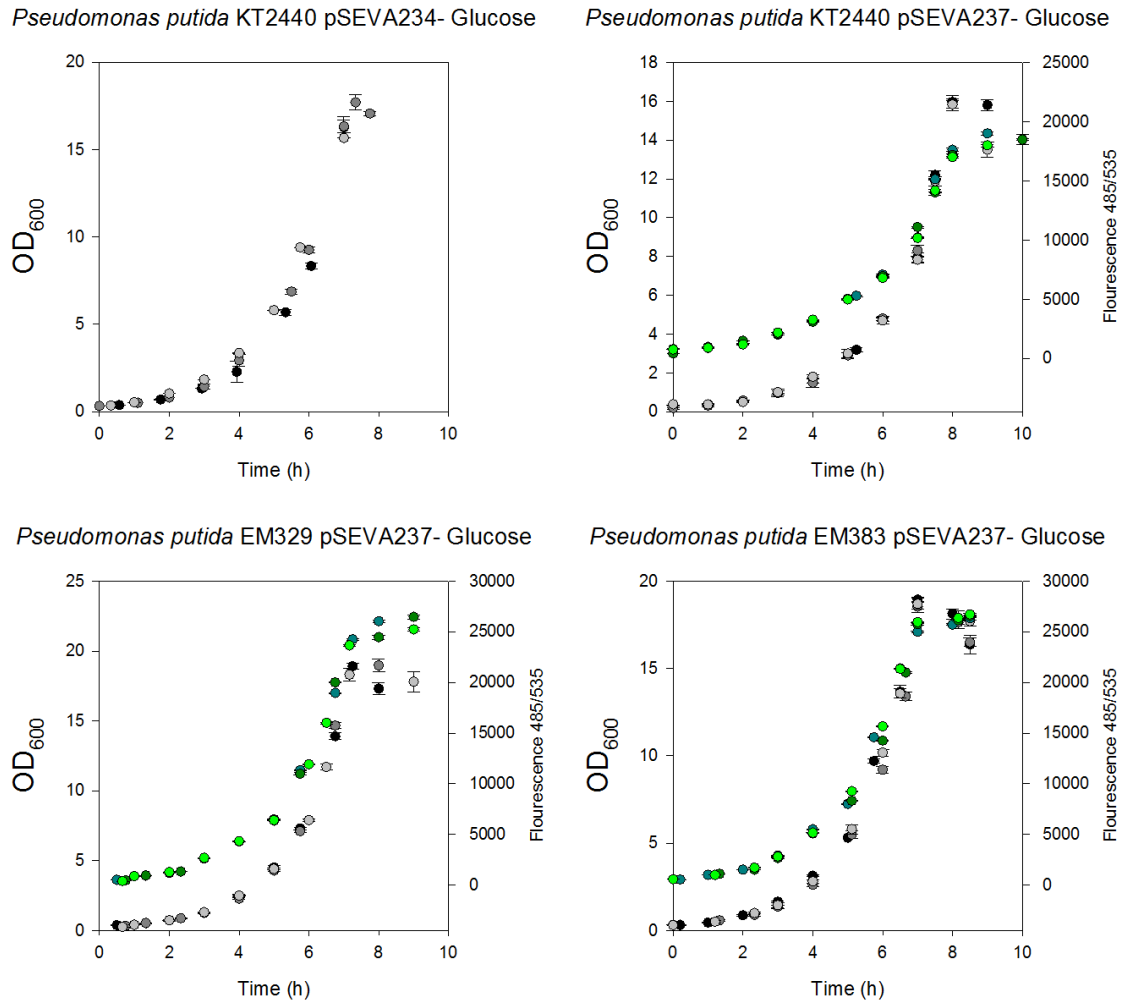


Figure C.9.: Physiological characterization in bioreactor batch cultivations of the different strains carrying plasmids. Batch cultivations were carried out with glucose as sole carbon source in a working volume of 1.5 liter in biological triplicates. The time course of the cultivations was monitored via biomass concentration (CDW; black, grey, and light grey dots) and in the case of the strains carrying GFP on the plasmid (pSEVA234G), the GFP fluorescence [GFP, measured in arbitrary fluorescence units (A.F.U.), dark green, green, and light green dots] was measured throughout the cultivation.

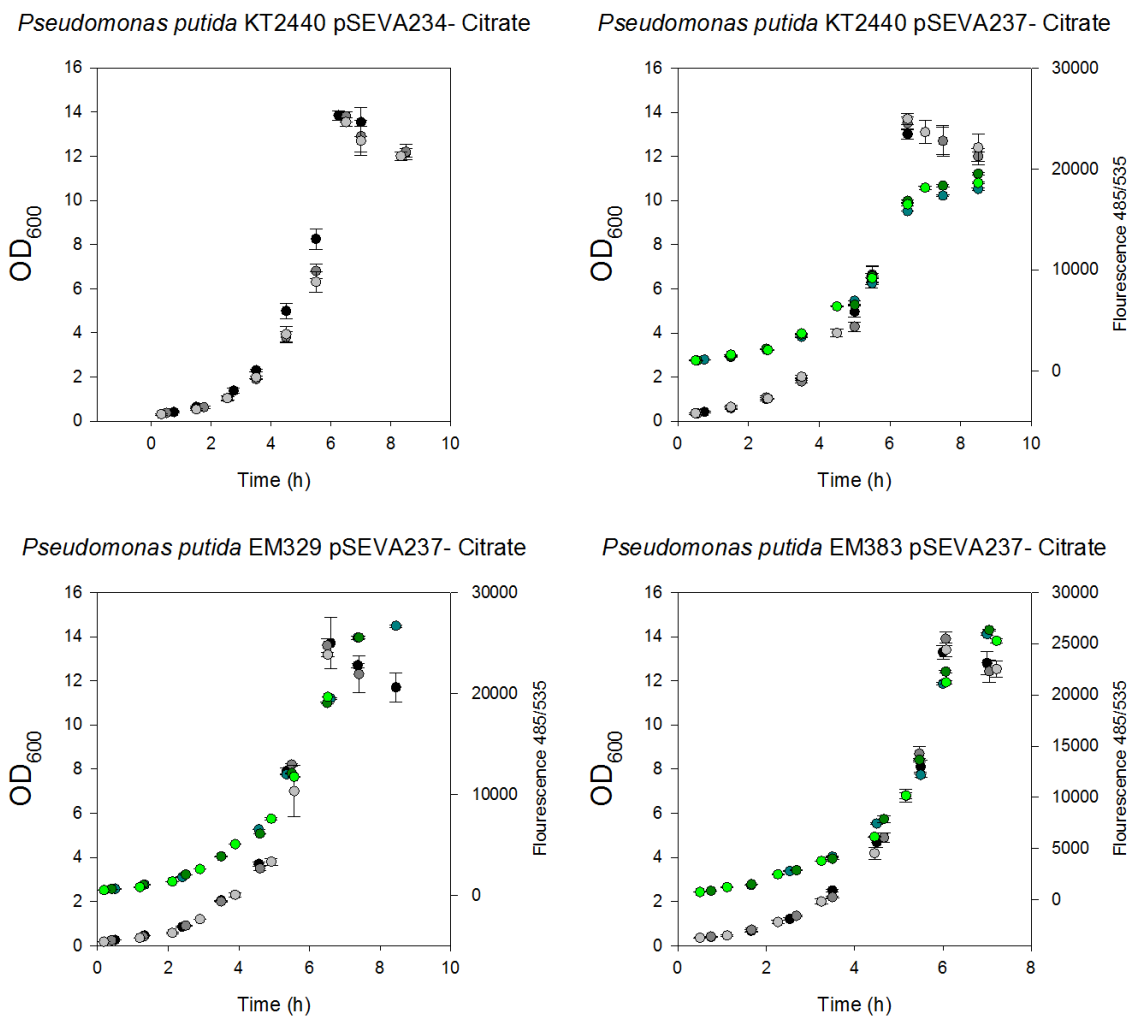


Figure C.10.: Physiological characterization in bioreactor batch cultivations of the different strains carrying plasmids. Batch cultivations were carried out with citrate as sole carbon source in a working volume of 1.5 liter in biological triplicates. The time course of the cultivations was monitored via biomass concentration (CDW; black, grey, and light grey dots) and in the case of the strains carrying GFP on the plasmid (pSEVA234G), the GFP fluorescence [GFP, measured in arbitrary fluorescence units (A.F.U.), dark green, green, and light green dots] was measured throughout the cultivation.