UNIVERSITÄT LEIPZIG

Characterization of population heterogeneity in a model biotechnological process using *Pseudomonas putida*

Von der Fakultät für Biowissenschaften, Pharmazie und Psychologie
der Universität Leipzig
genehmigte

DISSERTATION

zur Erlangung des akademischen Grades
Doctor rerum naturalium
(Dr. rer. nat.)

vorgelegt
von
Diplom-Biologe **Michael Jahn**
geboren am 31. Dezember 1985 in Dresden

Dekan: Prof. Dr. Erich Schröger
Gutachter: Prof. Dr. Hauke Harms
                   Prof. Dr. Susann Müller
                   Prof. Dr. Lars Blank

Tag der Verteidigung: 10. Juli 2015

Bibliographische Darstellung

Michael Jahn

Characterization of population heterogeneity in a model biotechnological process using *Pseudomonas putida*

Biotechnological processes are distinguished from classical chemistry by employing biomolecules or whole cells as the catalytic element, providing unique reaction mechanisms with unsurpassed specificity. Whole cells are the most versatile 'factories' for natural or nonnatural products, however, the conversion of e.g. hydrophobic substrates can quickly become cytotoxic. One host organism with the potential to handle such conditions is the gram-negative bacterium *Pseudomonas putida*, which distinguishes itself by solvent tolerance, metabolic flexibility, and genetic amenability. However, whole cell bioconversions are highly complex processes. A typical bottleneck compared to classical chemistry is lower yield and reproducibility owing to cell-to-cell variability. The intention of this work was therefore to characterize a model producer strain of *P. putida* KT2440 on the single cell level to identify nonproductive or impaired subpopulations. Flow cytometry was used in this work to discriminate subpopulations regarding DNA content or productivity, and further mass spectrometry or digital PCR was employed to reveal differences in protein composition or plasmid copy number.

Remarkably, productivity of the population was generally bimodally distributed comprising low and highly producing cells. When these two subpopulations were analyzed by mass spectrometry, only few metabolic changes but fundamental differences in stress related proteins were found. As the source for heterogeneity remained elusive, it was hypothesized that cell cycle state may be related to production capacity of the cells. However, subpopulations of one, two, or higher fold DNA content were virtually identical providing no clear hints for regulatory differences. On the quest for heterogeneity the loss of genetic information came into focus. A new work flow using digital PCR was created to determine the absolute number of DNA copies per cell and, finally, lack of expression could be attributed to loss of plasmid in non-producing cells. The average plasmid copy number was shown to be much lower than expected (~1 instead of 10-20). In conclusion, this work established techniques for the quantification of proteins and DNA in sorted subpopulations, and by these means provided a highly detailed picture of heterogeneity in a microbial population.

Contents

Summary

The chemical industry to date largely depends on fossil fuels which are further refined by classical organic chemistry. However, fossil fuels are a limited resource and the accompanying chemical reactions are energy-intensive and produce large amounts of waste, questioning overall sustainability of such processes. In contrast, biotechnological processes have the potential to substitute many classical chemical reactions and were identified as an important part of a sustainable economy in the future. Biotechnological processes stand out by employing biomolecules or whole cells as the catalytic element, providing unique reaction mechanisms with unsurpassed specificity, kinetics and sustainability. Processes based on biomolecules are generally driven under moderate temperature and pressure, take place in aqueous medium and produce mainly biodegradable waste. Whole cells are the most versatile 'factories' for natural or non-natural products, however, the conversion of troublesome substrates such as higher alcohols, aldehydes or aromatics quickly becomes cytotoxic even at low concentrations. One host organism with the potential to handle such conditions is the gram-negative soil-residing bacterium *Pseudomonas putida* (*P. putida*). Its compared to *Escherichia coli* one third larger genome comprises an impressive number of efflux pumps and other molecular mechanisms to remove deleterious substances from the cytoplasm. Unique remodeling of membrane lipids changing the membrane rigidity confers further resistance to solvents. It is equipped with many unusual catabolic routes leading to high metabolic flexibility, such as degradation of aromatic hydrocarbons like toluene or benzoate. *P. putida* is therefore a particularly suitable host for troublesome bioconversions involving hydrophobic solvents. However, whole cell bioconversions are often complex, not well-understood processes. Typical bottlenecks compared to classical chemistry are lower yield, difficult product purification and reduced reproducibility owing to cell-to-cell variability. The intention of this work was therefore to characterize a model producer strain of *P. putida* KT2440 on the single cell level to identify non-productive or impaired subpopulations. This strain was modified to produce a plasmid-encoded fusion protein of the target enzyme styrene monooxygenase and a variant of the green fluorescent protein as an optical reporter.

This enabled the investigation of *P. putida* on the single cell level using fluorescence microscopy and flow cytometry. Especially flow cytometry is capable of high-throughput analysis of thousands of single cells in a few seconds. Such optical methods are nevertheless restricted to

morphological analyses and it is difficult to obtain exact quantities of specific molecules. To gain a deeper insight on the molecular level, high throughput cell sorting was used to separate subpopulations, which were further investigated by label-free protein mass spectrometry. By this means, up to 1,000 out of 5,350 hypothetical proteins in *P. putida* could be quantified easily, providing a snapshot of the cellular proteome. A second downstream technique used in this work was digital PCR, which is able to quantify absolute copy numbers of DNA at very low concentration with high accuracy. It is not intended for simultaneous detection of many different molecule species, like it is the case for mass spectrometry, but for exact quantification of few specific DNA templates.

First of all, analysis of subpopulations by mass spectrometry required optimization of sample processing and storage conditions. It was unclear which sample preparation protocol was best to provide intact cells for sorting with little impact on proteome profile during storage. Three different methods, sodium azide fixation, deep freezing and vacuum drying were tested for durations of one week and one month. Deep freezing and vacuum drying were found to perform almost equally well and the former was used for further experiments. When productivity of the model *P. putida* strain was analyzed using the fluorescent reporter protein, it always showed a distinct bimodal distribution of low-producing and high-producing cells. These subpopulations were sorted and analyzed by protein mass spectrometry yielding a detailed picture of the gene expression landscape. Remarkably, only few changes were present in primary and secondary metabolism, but fundamental differences in stress related proteins. Obviously, producer cells tried to cope with protein stress only by 'damage control' – and no regulatory switch for the bimodal behavior was found. And, as a secondary result, all four plasmid-encoded proteins were almost absent in non-producing cells. Another common source of heterogeneity is cell cycling. It was hypothesized that cell cycle state (and consequently DNA content) are related to production capacity of cells. To harvest cells with different DNA content while all other parameters such as growth rate were kept stable, wild-type *P. putida* was grown in a chemostat and again analyzed by subpopulation proteomics. Interestingly, subpopulations of one fold, two fold, and higher DNA content were virtually identical providing no clear hints for regulatory differences in particular stages of cell cycling. However, a second finding was that growth rate tremendously influenced the gene expression program of *P. putida*. After transcriptional regulation was excluded as the reason for bimodality, the loss of genetic information came into focus. As previous results pointed towards a fundamental lack of expression of plasmid-encoded genes, a new work flow was created to determine the exact number of plasmid copies per cell. To this end, sorting of only 1,000 cells was combined with highly accurate digital PCR, which is theoretically able to detect a single DNA copy per reaction. And indeed, the average plasmid copy number was shown to be much lower than expected (~1 instead of 10-20 for 'medium' copy number). Plasmid copy number was highest (3-5) in producing cells while plasmids were completely absent in non-producing cells.

In conclusion, this work advanced the meth-

ods used for single cell analysis by optimizing sampling and storage of cells for 'subpopulation proteomics', a powerful combination of cell sorting and mass spectrometry. Furthermore, a new platform for cell sorting and digital PCR was developed to accurately measure absolute copy numbers of DNA. Besides methodological advances, it was shown that single-cell analysis provides a much more detailed picture of a microbial population in terms of productivity, cell cycle state or stress level. In particular, it was demonstrated that the population of *P. putida* KT2440 was clearly divided in groups of producing and non-producing cells, and that the ultimate molecular cause for this heterogeneity was plasmid loss. The producing cells were furthermore burdened by a considerable amount of stress caused by the heterologous gene expression. This can in the end confer a growth advantage to the non-producing plasmid-free subpopulation. In future applications, this source of heterogeneity might be eliminated using standardized plasmids with higher copy number or by engineering stable integration of target genes into *Pseudomonas*. Altogether, these findings would remain hidden by using standard bulk measurements, where cell-to-cell differences are averaged. It can nevertheless be expected that the progress in single cell analysis will unveil further and more delicate layers of microbial heterogeneity.

Zusammenfassung

Die chemische Industrie ist heutzutage zwingend auf fossile Rohstoffe angewiesen, welche mit klassischen organo-chemischen Prozessen weiterverarbeitet werden. Fossile Rohstoffe sind aber eine endliche Ressource und die zu ihrer Weiterverarbeitung genutzten Prozesse sind oftmals energieintensiv und verursachen giftige Abfälle, was die Nachhaltigkeit solcher Prozesse insgesamt in Frage stellt. Dem gegenüber stehen biotechnologische Prozesse, welche das Potential haben, viele konventionelle chemische Reaktionen zu ersetzen und als ein wichtiger Teil einer zukünftigen nachhaltigen „Bio-Ökonomie" gesehen werden. Biotechnologische Prozesse zeichnen sich aus durch die Verwendung von Biomolekülen oder ganzen Zellen als katalytische Elemente, welche eine Fülle einzigartiger Reaktionsmechanismen mit unerreichter Spezifität, Kinetik und Nachhaltigkeit mit sich bringen. Prozesse auf der Basis biologischer Katalysatoren spielen sich im Allgemeinen unter moderaten Druck- und Temperaturverhältnissen in wässrigen Medien ab, und hinterlassen vorwiegend biologisch abbaubare Abfallstoffe.

Ganze Zellen gelten als die vielseitigsten „Fabriken" für natürliche und nicht-natürliche Produkte. Allerdings kann zum Beispiel die Umwandlung von höherwertigen Alkoholen, Aldehyden oder Aromaten selbst bei geringen Konzentrationen schnell zu zytotoxischen Effekten führen. Ein Bakterium, welches potentiell besser mit solchen Bedingungen umgehen kann ist der gram-negative Bodenbewohner *Pseudomonas putida* (*P. putida*). Sein verglichen mit *Escherichia coli* um ein Drittel größeres Genom beinhaltet eine beeindruckende Anzahl an Genen für Efflux-Pumpen und andere molekulare Mechanismen, um schädliche Stoffe aus dem Zytoplasma zu entfernen. Der einzigartige Mechanismus der Modifikation von Membranlipiden verleiht der Zellmembran eine veränderte Festigkeit und führt zu größerer Lösungsmitteltoleranz. Es ist außerdem ausgestattet mit etlichen ungewöhnlichen katabolischen Wegen die zu einer hohen metabolischen Flexibilität führen, zum Beispiel der Nutzung von aromatischen Kohlenwasserstoffen wie Toluol oder Benzoesäure. *P. putida* ist daher ein besonders geeigneter Organismus zur Ausführung schwieriger Biokonversionen, bei denen zum Beispiel hydrophobe Lösungsmittel beteiligt sind. Dennoch sind auch Ganzzell-basierte Bioverfahren komplexe Prozesse und bisher noch unzureichend verstanden. Typische Engpässe und Einschränkungen verglichen mit herkömmlichen chemischen Verfahren sind eine geringere Ausbeute, schwierige Aufarbeitung des Produkts und geringere Reproduzierbarkeit durch interzelluläre Variabilität. Das Ziel dieser Arbeit war da-

her, den Modellorganismus *P. putida* KT2440 auf der Ebene der einzelnen Zelle zu charakterisieren, um schlecht produzierende oder geschädigte Subpopulationen während eines Produktionsprozesses zu identifizieren. Dafür wurde ein Stamm verwendet, der ein Plasmid-kodiertes Fusionsprotein herstellt, welches aus dem Enzym Styrol-Monooxygenase und einem grün fluoreszierenden Protein als optischer Markierung besteht.

Diese Vorgehensweise erlaubte die Untersuchung von *P. putida* auf der Einzelzell-Ebene mittels Fluoreszenz-Mikroskopie und Durchfluss-Zytometrie. Besonders Letztere ist als Hochdurchsatz-Technik in der Lage, tausende einzelne Zellen in wenigen Sekunden zu vermessen. Derartige optische Verfahren sind aber begrenzt auf rein morphologische Analysen und es ist schwierig, absolute Mengen bestimmter Moleküle nachzuweisen. Um dennoch einen tieferen Einblick auf der molekularen Ebene zu bekommen, wurde Hochdurchsatz-Zellsortierung eingesetzt, um Zellen verschiedener Subpopulationen zu isolieren. Diese wurden weitergehend mittels markierungsfreier Protein-Massenspektrometrie (MS) analysiert. Mit dieser Technik konnten bis zu 1000 von insgesamt 5350 hypothetischen Proteinen in *P. putida* quantifiziert werden, und damit eine detaillierte Momentaufnahme des zellulären Proteoms erhalten werden. Eine zweite nachgeschaltete Technik, die im Rahmen dieser Arbeit eingesetzt wurde, war digitale Polymerase-Kettenreaktion (digitale PCR). Diese neuartige Technik erlaubt die absolute Quantifizierung von in geringsten Kopienzahlen vorhandenen DNA-Molekülen mit hoher Genauigkeit. Im Gegensatz zur MS, die eine hohe Anzahl verschiedener Proteine parallel nachweisen kann, ist digitale PCR für die exakte Quantifizierung weniger bestimmter Moleküle vorgesehen. Zunächst einmal erforderte die Analyse von Subpopulationen mittels MS eine gründliche Optimierung der Probenahme und -Lagerung. Es war bisher unklar, welches Probenahme-Protokoll die besten Ergebnisse liefern könnte im Hinblick auf Intaktheit der Zellen für die Sortierung bei gleichzeitig geringem Einfluss auf das Proteom-Profil. Dafür wurden drei gänzlich verschiedene Methoden für eine Lagerungsdauer von einer Woche und einem Monat getestet: Natriumazid-Fixierung, Tiefkühlung, und Vakuumtrocknung. Tiefkühlung und Vakuumtrockung erzielten letztendlich die besten Ergebnisse und erstere Methode wurde für alle weiteren Experimente benutzt. In den folgenden Versuchen wurde die Produktivität des Modellstammes *P. putida* mit Hilfe des grün fluoreszierenden Proteins untersucht, und es wurde eine interessante bimodale Aufteilung in starkproduzierende und wenig-produzierende Zellen festgestellt. Um die Ursache für diese Heterogenität zu finden wurden beide Subpopulationen sortiert und mittels MS analysiert. Dies ergab einen detaillierten Einblick in die gesamte Genexpression der Zellen. Bemerkenswerterweise wurden nur wenige Unterschiede im primären und sekundären Stoffwechsel gefunden, während vor allem stressassoziierte Proteine stark reguliert waren. Offensichtlich versuchten die Produzenten, dem durch übermäßige Proteinproduktion enstandenen Stress nur durch Schadensbegrenzung zu begegnen, während weder eine Umlenkung von Kohlenstoffflüssen noch Anzeichen für einen regulatorischen Schalter gefunden wurde, der die Un-

terschiede in der Produktivität erklärt hätte. Einen Hinweis lieferte nur die Tatsache, dass alle vier Plasmid-kodierten Proteine in nicht-produzierenden Zellen insgesamt kaum vorhanden waren. Eine weitere Quelle von Heterogenität ist der Ablauf des Zellzyklus. Es wurde angenommen, dass auch der Zellzyklus (und damit der DNA-Gehalt) einen Einfluss auf die Produktionskapazität der Zellen hat. Um Zellen verschiedenen DNA-Gehalts bei gleichzeitiger Konstanz aller anderen Parameter wie z. B. Wachstumsrate zu erhalten, wurde ein Wildtyp-Stamm von *P. putida* in Chemostaten kultiviert und die Subpopulationen wiederum mittels MS analysiert. Interessanterweise konnte kein wesentlicher Unterschied zwischen Zellen mit einfachem, zweifachem, oder höherem DNA-Gehalt festgestellt werden, und damit auch kein klarer Hinweis auf regulatorische „Programme" während des Zellzyklus. Allerdings wurde eine Veränderung der Wachstumsrate als Schlüsselereignis für das Genexpressionsmuster der Zellen erkannt. Nachdem transkriptionelle Regulation weitestgehend als Ursache für Bimodalität ausgeschlossen werden konnte, rückte der Verlust genetischer Information in den Fokus. Vorangegangene Ergebnisse zeigten einen Mangel plasmidkodierter Proteine in Nichtproduzenten, daher wurde ein Workflow entwickelt, um die exakte Plasmidkopienzahl pro Zelle zu bestimmen. Dafür wurde die übliche Sortierung auf nur 1000 Zellen herunterskaliert und diese mittels neuartiger digitaler PCR kombiniert, welche theoretisch in der Lage ist, einzelne DNA-Moleküle in einer Standard-PCR-Reaktion zu detektieren. Und tatsächlich war die durchschnittliche Plasmidkopienzahl wesentlich geringer als erwartet (~1 statt 10-20 bei „mittler-

er" Kopienzahl). Die Kopienzahl war außerdem deutlich höher in produzierenden Zellen (3-5), während nicht-produzierende Zellen praktisch plasmidfrei waren.

Insgesamt konnte diese Arbeit wesentlich zum methodischen Fortschritt der „Subpopulations-Proteomik" beitragen, indem Probenahme und -Lagerung von ganzen Zellen optimiert wurden. Zudem wurde eine neue Plattform für die Sortierung kleinster Zellzahlen und anschließender Bestimmung der DNA-Kopienzahl mittels digitaler PCR etabliert. Abgesehen von methodischen Verbesserungen konnte gezeigt werden, dass mit Einzelzellanalyse ein wesentlich detaillierteres Bild einer mikrobiellen Population erhalten werden kann, was z. B. Produktivität, Zellzyklus-Stadium oder Stresslevel angeht. Im Besonderen zeigte die Population von *P. putida* KT2440 eine klare Zweiteilung in produzierende und nicht-produzierende Zellen, die letztendlich auf Plasmidverlust zurückgeführt werden konnte. Heterologe Genexpression führte außerdem zu beachtlichem Stress in produzierenden Zellen und verleiht nicht-produzierenden auf Dauer einen Wachstumsvorteil. In zukünftigen Anwendungen kann diese Art genetischer Heterogenität zum Beispiel durch Verwendung standardisierter Plasmide mit höherer Kopienzahl oder stabilen genomischen Integrationen ausgeschlossen werden. All diese Erkenntnisse wären verborgen geblieben durch übliche, auf der Populationsebene durchgeführte Messungen, die alle Unterschiede zwischen einzelnen Zellen verwischen. Es ist außerdem zu erwarten, dass weitere technische Sprünge in der Einzelzellanalytik einen noch reicheren Fundus an mikrobieller Heterogenität freilegen werden.

1 Introduction

1.1 Biotechnology for a new economy

Until today, the overwhelming majority of industrial processes for the synthesis of chemicals is based on petrochemistry. However, natural fossil resources such as crude oil are getting more and more scarce and the extensive use of oil-based products has led to severe pollution of the environment. In the past years, biotechnology has made great promises as an alternative to chemical synthesis, aiming at the replacement of energy-intensive and non-sustainable processes by biotransformation (Figure 1.1). Such processes are distinct from traditional chemical synthesis by a number of features. Bioprocesses are driven by whole cells or cellular biomolecules such as enzymes or ribozymes. These catalysts are self-replicating, inexpensive, non-toxic and biodegradable. This is opposed to chemical catalysts, which often comprise rare and expensive elements such as platinum. Furthermore, enzymatic reactions usually take place at moderate temperature and pressure, and are highly specific for their substrate. Due to the unique reaction mechanisms, the enantiomeric selectivity can be up to 100 %.

However, many issues still prevent a major breakthrough of biotechnological processes. The most important drawback is the complexity of whole cells or biomolecules, reducing the predictability and reproducibility of bioprocesses. In case of whole cells, the genome ususally encodes at least several hundred different proteins per cell (Tsoy *et al.*, 2013), which interact in a highly complex way to keep the cellular homeostasis. And even purified catalytic biomolecules such as enzymes are in turn composed of many different amino acids, contributing to a flexible and sometimes delicate conformation. Changes in protein conformation can lead to reduced or altered reactivity. In addition, biomolecules are rapidly degraded under harsh environmental conditions such as extreme pH, temperatures or reactive chemicals, which is not the case for elemental catalysts.

1.2 Whole cells as biocatalysts

Traditionally, microbial cells have been used as biocatalysts for hundreds of years in human history. Famous examples are *Saccharomyces* yeasts for brewing beer or fermentation of other alcoholic beverages, lactic acid bacteria for fermentation of dairy products, or microbial consortia for

Figure 1.1: Scheme of the 'bio-economy', a concept of sustainable product manufacturing based on renewable resources and biocatalytic processes. Figure is reproduced from Biobased-Industries-Consortium (2015).

preparation of sour dough. These processes utilize naturally available carbon and energy sources to yield alcohols or organic acids as useful by-products of the microbial metabolism. However, much more challenging are whole cell-driven bioprocesses where non-natural or even toxic substances are the desired reaction substrate or product. Many microbial species were already found to naturally metabolize otherwise toxic compounds by using these as a carbon or energy source, although in small quantities. Among these substrates are n-alkanes (Smits *et al.*, 1999) as well as aromatic hydrocarbons like toluene, benzoate or styrene (Panke *et al.*, 1998; Udaondo *et al.*, 2012) or even antibiotics (Lin *et al.*, 2015). However, although the bioconversion of small amounts of these compounds by microbial cells can be implemented easily, a biotechnological process requires many more fulfilled conditions to be feasible and cost-efficient:

- High yield

- High productivity (volumetric or time-dependent)

- Growth on cheap/widely available carbon source

- Growth on complex substrates (e.g. 2nd generation biomass)

- Tolerance for reaction substrate and product

- Easy downstream processing

- Stability and re-use of the catalytic unit

- General process stability and reproducibility

1.3 Engineering microbial 'cell factories'

1.3.1 Optimizing recombinant gene expression

The aforementioned requirements can be seen as guide lines for the construction of microbial 'cell factories', cells that have been deliberately and systematically engineered for a specific purpose. In the past decades a range of different organisms and strains turned out to be suitable hosts for different tasks. Traditionally, *Escherichia coli* (*E. coli*) has been the workhorse for the production of all kinds of proteins, or the subsequent enzymatic conversion of a biochemical compound. This is owing to the important historical role *E. coli* has played as a model organism in molecular genetics, contributing heavily to the understanding of gene structure and regulation, or the fundamental processes of transcription and translation (Cooper, 2000). Moreover, it was one of the first organisms having a fully sequenced genome made publicly available in 1997 (Blattner *et al.*, 1997), enabling researchers to target specific genes for knockout, modification or over-expression. For long time, researchers and companies have resorted to *E. coli* as a production host by optimizing parameters such as cultivation media, expression strains, vectors and growth conditions using a trial and error approach (Makino *et al.*, 2011). Meanwhile, ever-growing knowledge paved the way for further improvements of bioprocesses. Examples are codon-optimization where the amino acid encoding base pair triplets are exchanged for ones heavily used by the host organism, or the co-expression of so called 'chaperones', proteins assisting and governing polypeptide maturation (Makino *et al.*, 2011).

1.3.2 Systems biology and synthetic biology

More recently, the advance of bioinformatics opened the new field of 'synthetic biology'. According to a definition by Serrano (2007) synthetic biology is the rational design and implementation of new-to-nature functions in biological systems. Key characteristics are the adaptation of an 'engineering perspective', standardization and modularization of implemented functions to allow recombination and re-use, and the broad use of bioinformatics for rational design. Synthetic biology is closely related to 'systems biology', meaning the analysis of complete biological systems instead of an isolated perspective on single genes or proteins, but synthetic biology strongly emphasizes the design of *novel* functions (Serrano, 2007). A major breakthrough for synthetic biology, besides the rapid advances in bioinformatics, was the opening of a new molecular biology toolbox for deleting, modifying or adding new genes in microbes. These new tools broadly cover different aspects of molecular biology, such as *in vitro* modification of DNA (facilitated cloning, gene synthesis), *in vivo* modification of genomic DNA (genome editing), or techniques for efficient screening of clone libraries. Seamless cloning, for instance, is a technique obviating the need for traditional restriction and ligation steps. It is based on PCR amplification of all required DNA

fragments followed by direct assembly in one tube without repetitive cloning steps (Gibson *et al.*, 2009). This technique was used for the complete assembly of an *in vitro* synthesized genome of *Mycoplasma mycoides*, representing the first cell controlled by a fully artificial genome (Gibson *et al.*, 2010).

1.3.3 Genetic engineering of *Pseudomonas putida*

Other techniques were developed to implement the genes for novel metabolic pathways into the microbial genome. Originally, genome modification was restricted to non-targeted methods such as successive rounds of mutagenesis by chemical treatment or UV ray and selection for desired mutants. Meanwhile, alternative techniques allow a much more targeted approach to modify the genome. By employing recombination systems, which can be found in many organism from bacteriophages to eukaryotes, it is possible to introduce a DNA sequence of choice at a desired locus in the genome. To this end, the prepared template DNA must be equipped with sequences homologous to the integration site at both termini and will integrate upon co-expression of a recombinase, as first demonstrated in *E. coli* (Datsenko & Wanner, 2000). After integration and selection, antibiotic resistance genes can be removed from the genome by another round of recombination, e.g. based on FLP recombinase targeting specific *frt* recognition sites. This way, a complete library of around 4,000 single-gene knockout strains of *E. coli* was created termed the 'Keio collection' (Baba *et al.*, 2006).

An interesting host for biotechnology is the soil bacterium *Pseudomonas putida* (*P. putida*). Besides its metabolic versatility (section 1.4.1), genetic engineering to e.g. introduce new genes and pathways is straightforward in *P. putida*. The Tn5-derived mini-transposon system is the method of choice for genomic DNA integration and was for years successfully applied for repetitive insertions of DNA into the same cell (de Lorenzo *et al.*, 1998; Poblete-Castro *et al.*, 2012). A major step was made recently by introducing a new genome editing method, allowing the *in situ* insertion or deletion of long DNA sequences Martínez-García & de Lorenzo (2011, 2012). It is a two-step process with the first step being the integration of a suicide vector into the host genome via two flanking homologous sequences; The original genomic DNA residing between the two flanks is removed. In a second step, the restriction enzyme *I-Sce*1 specifically cuts the inserted vector DNA. The DNA break will be repaired by the cellular recombination machinery fusing the open ends and eventually removing the original vector DNA. These recombination-based systems have the great advantage of not leaving an antibiotic marker in the genome. This allows the recycling of one and the same antibiotic resistance in multiple sequential deletions. As an example, Martínez-García & de Lorenzo (2011) used *I-Sce*1 recombination to eradicate the large operons *ttgABC* and *mexEF/oprN* responsible for pumping out antibiotics, and suspected prophage regions of up to 40 kbp length.

1.3.4 Standardization of biological 'parts'

The tailoring of microbes demands not only the precise deletion of DNA but also the insertion of external DNA encoding novel functions, such as gene expression cassettes, regulatory networks or whole metabolic pathways. New constructs are constantly assembled by scientists and tested in various organisms, however, many of them are often not transferable from one expression host to another due to the lack of standardization (Serrano, 2007). To overcome this limitation, MIT-based researchers developed a standardization guideline (Knight, 2003) and launched a public repository called the 'Registry of Standard Biological Parts' (*http://parts.igem.org*), together with an international contest on genetically engineered machines (iGEM). These parts called 'BioBricks' comprise basic genetic elements such as promoters, terminators or reporter genes or more complex assemblies thereof. By these means highly elaborate modules were assembled, such as *E. coli* strains with a scent of banana (IGEM, 2006), an engineered surface adherence by introduction of a synthetic *curli* operon (Drogue *et al.*, 2012), or a Pavlovian conditioning circuit (Zhang *et al.*, 2014).

A similar rationale is behind the foundation of the **S**tandard **E**uropean **V**ector **A**rchitecture (SEVA) providing many parts for *Pseudomonas* (Silva-Rocha *et al.*, 2013). The SEVA principle is to construct, collect and openly distribute plasmids with a particular standardized structure. SEVA plasmids feature three major elements with different options for each to choose. These are an antibiotic resistance cassette, the replication origin, and a custom cargo site. All parts except the cargo are predefined allowing standardized cloning strategies, easy subcloning and better comparability in experiments.

1.4 *Pseudomonas* as microbial cell factory

1.4.1 Advantages of *Pseudomonas putida*

Pseudomonas putida is a ubiquitous gram-negative soil resident but also an interesting host for biotechnology. It is generally recognized as safe as it shares only 85 % sequence similarity with its famous human-pathogenic counterpart, *Pseudomonas aeruginosa*, and is known to lack typical virulence factors for host invasion (Nelson *et al.*, 2002). This is most likely owing to its different lifestyle as a rhyzosphere inhabitant rather than human parasite. *P. putida* is well-known for the degradation of otherwise toxic, ecologically harmful and resilient chemicals, especially aromatic hydrocarbons. Among those compounds are toluene (Díaz *et al.*, 2008), benzoate and xylene (Martínez-García & de Lorenzo, 2012). Interestingly, many genes with degradative purpose are encoded on plasmids such as the TOL plasmid ('pWW0' in *P. putida*), carrying the operon for toluene and xylene degradation (Williams & Murray, 1974; Greated *et al.*, 2002). Not taking

plasmids into account, the genome of the widely used model strain *P. putida* KT2440 contains 5,350 genes on 6.2 Mbp DNA, considerably more than e.g. *E. coli* with 4377 genes on 4.6 Mb (*pseudomonas.com*, Winsor *et al.* (2011)). Many of *P. putida*'s genes encode transporters (or parts of them), e.g. for exporting cytotoxic ions or for importing a large variety of sugars, amino acids or aromatic compounds as carbon sources. Altogether, 26 transporters for ions (chromium, arsenic, zinc and cobalt amongst others) were identified *in silico*, 13 transporters for carboxylated and aromatic coumpounds, and at least 12 efflux pumps for drugs or unknown chemicals (Nelson *et al.*, 2002). In the microbial world *P. putida* possesses more transporters for aromatics than any other sequenced organism (Nelson *et al.*, 2002). Its metabolic versatility acquired during evolution (and highlighted in the next section) makes it therefore a promising host for biotechnology.

1.4.2 Carbon utilization of *Pseudomonas putida*

Although *P. putida* is a versatile organism, it is a complex organism as well. Some crucial differences to other commonly used host organisms must be noted. In the first place the central carbon metabolism of *P. putida* differs to other organisms by its utilization of many different carbon sources, while glucose is explicitly not the preferred carbon source, but organic acids such as succinate, malate or citrate (Poblete-Castro *et al.*, 2012). Although *P. putida* obviously prefers other carbon sources for hexoses, it still has three pathways to metabolize glucose (Figure 1.2), the Embden-Meyerhoff-Parnas pathway (EMP, 'glycolysis'), the Entner-Douderoff pathway (ED) and the pentose phosphate pathway (PPP) (Poblete-Castro *et al.*, 2012). Distinct from most other bacteria *P. putida* channels 96 % of glucose through the ED pathway and cannot use glycolysis, due to the lack of phosphofructokinase (*pfk*, Figure 1.2) (Chavarría *et al.*, 2012). The PPP pathway is mainly used for anabolic reactions providing building blocks for cell maintenance. If not for glucose, what other carbon sources are suitable for *Pseudomonas*? Naturally not being able to grow on C5 sugars, efforts were successful to engineer *P. putida* S12 for growth on the C5 sugars xylose and arabinose by heterologous expression of the respective genes from *E. coli* and *Caulobacter crescentus* (Meijnen *et al.*, 2008, 2009). An interesting feedstock which can be metabolized by *P. putida* is the C3 sugar alcohol glycerol as it is a common by-product in the biodiesel industry (Poblete-Castro *et al.*, 2012). It has been used as sole carbon source for the lab scale production of such diverse chemicals as medium chain length PHA or geranic acid (Figure 1.3, Poblete-Castro *et al.* (2014); Mi *et al.* (2014)). However, organic acids directly fueling the TCA are the preferred carbon source of *P. putida*, which corresponds well to its lifestyle as a soil bacterium (Nelson *et al.*, 2002). In soil, metabolic intermediates like organic acids are thought to be more common than sugars. Other atypical organic acids that can be metabolized are butyrate (Cerrone *et al.*, 2014) or benzoate (Mandalakis *et al.*, 2013).
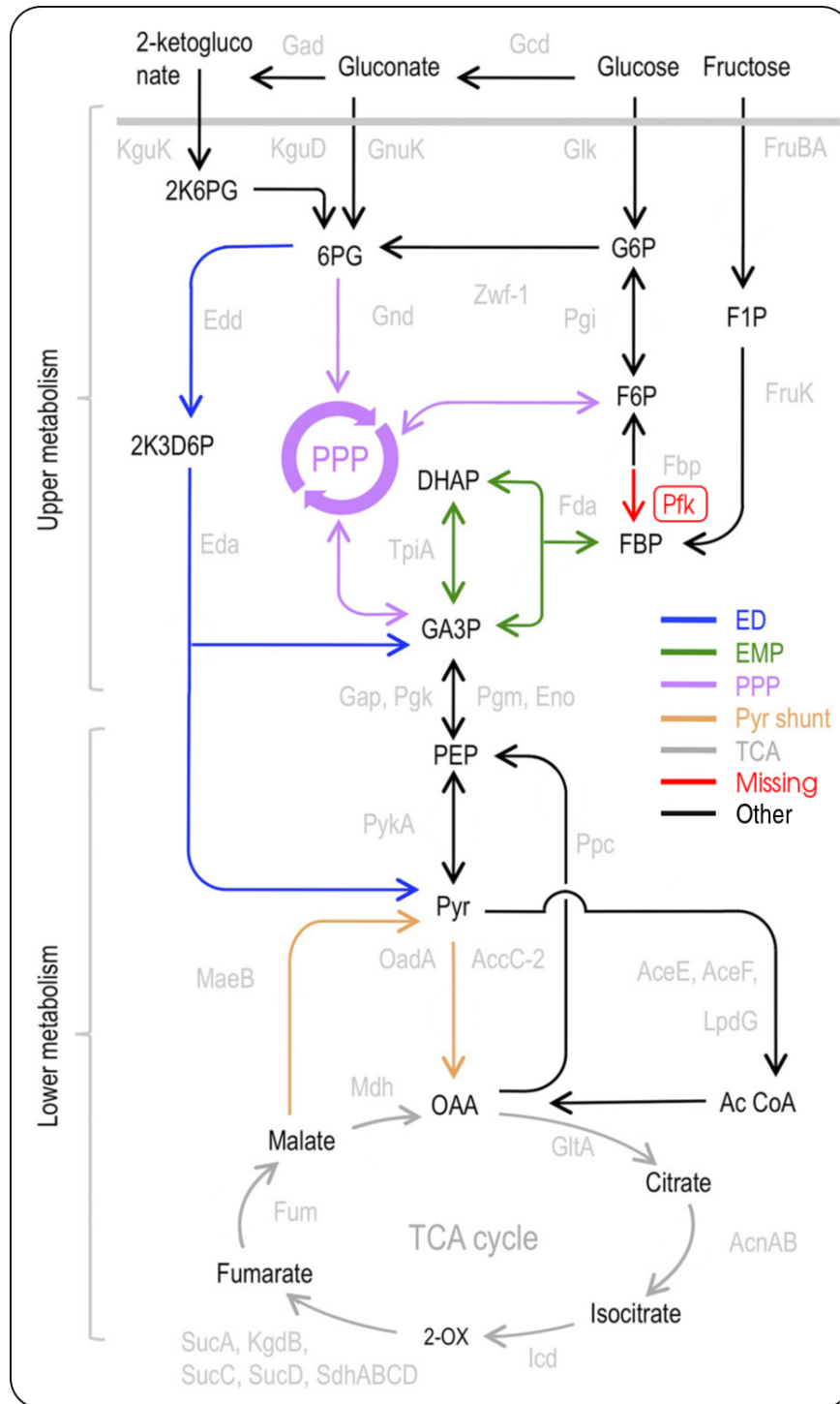
Figure 1.2: The central carbon metabolism of *Pseudomonas putida* is displayed as divided into an upper and lower part. The upper part includes uptake of substrates such as fructose or glucose and metabolization via three different pathways, the Embden-Meyerhoff-Parnas pathway (EMP or 'glycolysis'), the Entner-Douderoff pathway (ED) and the pentose phosphate pathway (PPP). Note that the phosphofructokinase (Pfk, red) required for glycolytic comsumption of glucose is not functional in *P. putida*. Yellow, pyruvate shunt. Grey, tricarboxylic acid cycle. Figure is modified from Chavarría *et al.* (2012).
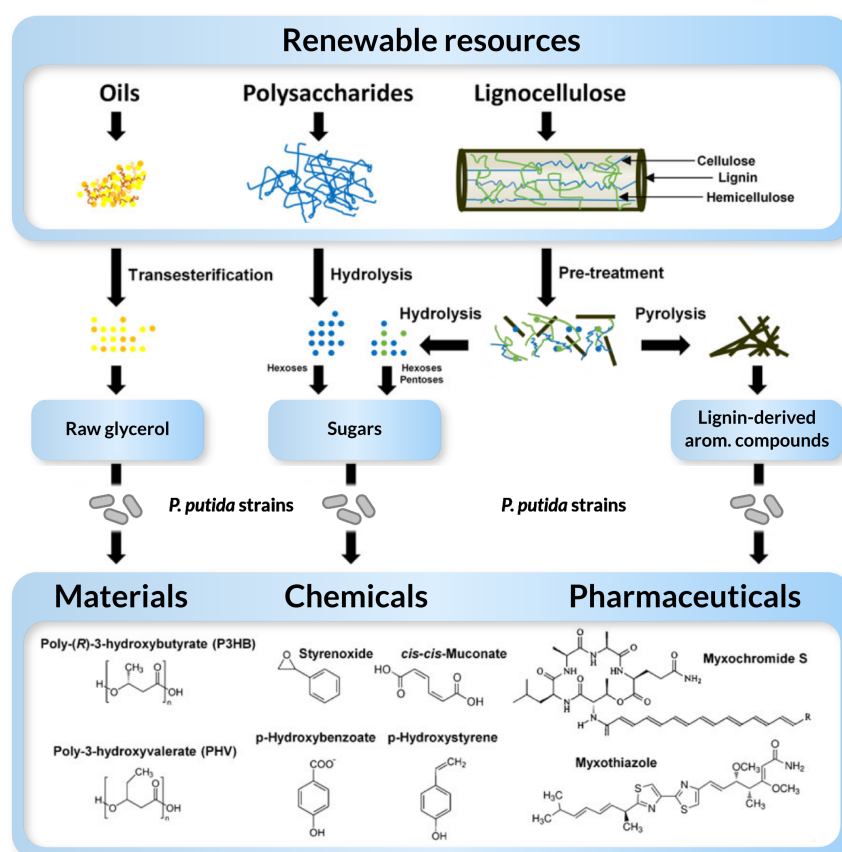
Figure 1.3: Applications of *Pseudomonas putida* in industrial biocatalysis. Figure is modified from Poblete-Castro *et al.* (2012).

1.4.3 Mechanisms of solvent tolerance in *Pseudomonas putida*

Pseudomonas putida is an environmental generalist and, as outlined before, has a highly versatile metabolism. Although the full genome sequence of many strains is available by now, the metabolism and its regulation are not as well understood as it is the case for *E. coli*. While the lack of knowledge and the smaller arsenal of genetic tools is clearly a drawback, the major strength of *Pseudomonas* strains is solvent tolerance, enabling reactions with organic solvents as reactants, products or as a second phase for scavenging the target compounds (Blank *et al.*, 2008; Rühl *et al.*, 2009). Different mechanisms are known to contribute to the phenotype of solvent tolerance (Figure 1.4).

First, *P. putida* is able to adjust the lipid composition of the cytoplasmic membrane, as solvents are preferably deposited in the membrane, thereby increasing fluidity in a harmful way (Segura *et al.*, 2012). To counteract that, cells enzymatically switch *cis*-configured unsaturated fatty acids to the more compact *trans*-configuration, rendering the membrane more rigid (Heipieper *et al.*, 2003). In a similar way, cells can change the ratio of saturated and unsaturated fatty acids
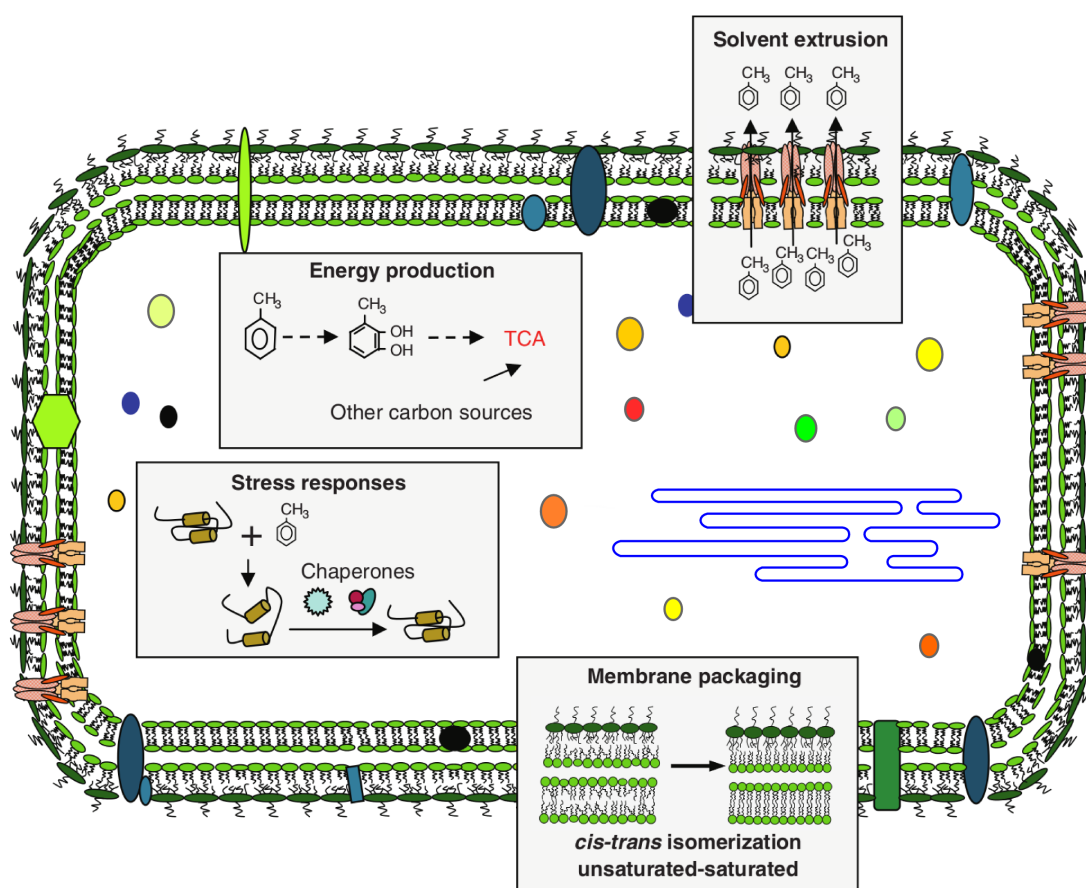
Figure 1.4: The four mechanisms of solvent tolerance in *Pseudomonas*. Figure is modified from Segura *et al.* (2012).

(Segura *et al.*, 2012). A novel membrane-located mechanism is the formation and release of extracellular membrane vesicles which was shown to increase membrane hydrophobicity and therefore endurance against chemical stresses (Baumgarten *et al.*, 2012). Second, solvent efflux pumps are able to remove toxic compounds from the cytoplasm, making some *Pseudomonas* strains with a higher number of pumps like DOT-T1E (Molina *et al.*, 2011; Udaondo *et al.*, 2012) or S12 (Tao *et al.*, 2012) even more solvent-tolerant than e.g. KT2440. In DOT-T1E, three operons (*ttgABC, ttgDEF, ttgGHI*) are responsible for resistance against toluene and presumably other solvents. Solvent tolerance is an inducible phenotype and traditionally requires cumbersome physiological adaptation to be fully expressed. However, knocking out the main regulator *ttgV* of the efflux pump encoding operon *ttgGHI* was shown to establish solvent tolerance immediately (Volmer *et al.*, 2014). Third, physiological counter measures can include the activation of enzymes responsible for removal of reactive oxygen species as a by-product of detoxification (Choi *et al.*, 2014), formation of trehalose as a protective agent, expression of chaperones to improve protein folding (Simon *et al.*, 2014), or increasing intracellular concentration of cofactors (Udaondo *et al.*, 2012). *Pseudomonas* has a higher regeneration rate of cofactors and therefore higher

buffer capacity than other common bacteria (Blank *et al.*, 2008; Ebert *et al.*, 2011). And fourth, Pseudomonads are able to enzymatically degrade many toxic compounds and use them as carbon source as outlined in section 1.4.1. Of all the described mechanisms, solvent efflux pumps are regarded as the dominant means of conferring solvent tolerance (Segura *et al.*, 2012).

1.4.4 Processes based on *Pseudomonas putida* to date

Despite the many known advantages of *Pseudomonas* as a host organism for biotechnology, only a handful of processes were actually implemented by the chemical industry. According to Poblete-Castro *et al.* (2012), 22 production processes with industrial relevance were already outlined and tested for *P. putida*. Among them are many fine chemicals but also different 'flavors' of polyhydroxy alkanoic acid (PHA) as a raw material for future biomaterials. In contrast to that, only nine processes were actually implemented by the industry, the majority of them being pharmaceuticals (Table 1.1). Nevertheless it can be assumed that the market for biotechnological production of platform chemicals will increase in the near future, and *Pseudomonas* is a promising candidate for those reactions involving hydrophobic compounds.

Table 1.1: Implemented biotechnological processes using *Pseudomonas* strains. Table is reproduced from Poblete-Castro *et al.* (2012).

Product	Biocatalyst	Applicability	Company
2-Quinoxalinecarboxylic acid	*P. putida* ATCC 33015	Biological activity	Pfizer (US)
5-Methylpirazine-2-carboxylic acid	*P. putida* ATCC 33015	Pharmaceutical	Lonza (SUI)
Chiral amines	*Pseudomonas* sp. DSM 8246	Biological activity	BASF (GER)
5-Cyanopentanamide	*P. putida*	Catalysis	DuPont (US)
(S)-2-Chloropropionoc acid	*Pseudomonas* sp.	Herbicides	Astra Zeneca (US)
D -p-Hydroxyphenyl glycine	*P. putida*	Pharmaceutical	Several companies
Chiral compounds	*P. putida* ATCC 12633	Pharmaceutical	DSM (NL)
4-(6-Hydroxypyridin-3-yl)-4-oxobutyrate	*Pseudomonas* sp. DSM 8653	Pharmaceutical	Lonza (SUI)
Paclitaxel	*Pseudomonas* sp. lipase AK	Pharmaceutical	Bristol-Myers Squibb

1.4.5 Styrene epoxidation as a model reaction

Styrene is a hydrophobic aromatic compound consisting of a C6 ring with an ethylene group. It is an important monomeric building block for the synthesis of polystyrene polymers ('styrofoam') and is produced at a rate of 25 million metric tons worldwide with a value of 1,300 USD per ton (Exelus, 2012). A commercially interesting reaction is the stereoselective epoxidation of styrene,
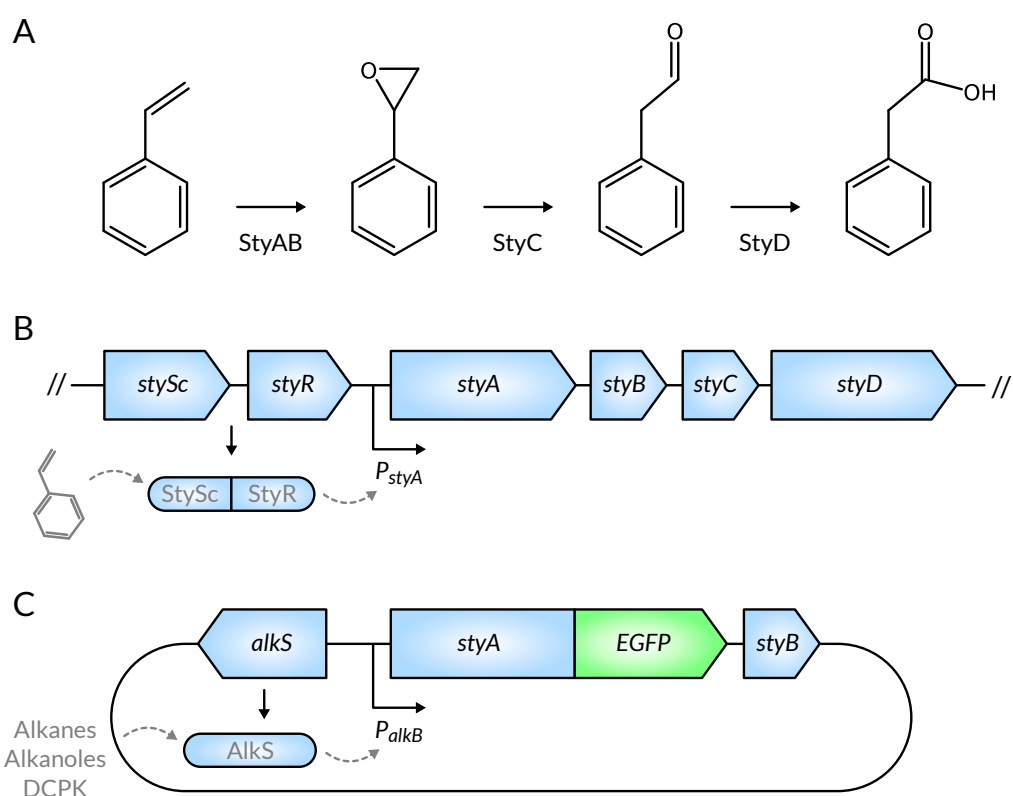
Figure 1.5: A) Pathway of enzymatic styrene degradation in *P. taiwanensis* VLB120 leading from styrene to styrene oxide to phenylacetic aldehyde and finally phenylacetic acid. The genetic structure of the respective operon is shown in B): *stySc*, encoding the styrene sensor component, *styR*, the transcriptional regulator, *styA* and *styB*, together encoding the styrene monooxygenase, *styC*, the styrene oxide isomerase, and *styD*, the dehydrogenase. C) The model expression system in this work are plasmids encoding a StyA-EGFP fusion protein and StyB. Transcription is regulated by AlkS acting on the *alkB* promoter (P_{alkB}). Figure modified from Panke *et al.* (1998).

forming a highly reactive oxygen-containing 3-heterocycle (Figure 1.5 A). The obtained styrene oxide can be used for further synthesis of fine chemicals or pharmaceuticals. The soil bacterium *Pseudomonas taiwanensis* VLB120 was shown to grow on styrene as a sole carbon and energy source and the respective degradative genes could be identified, isolated and characterized (Panke *et al.*, 1998). The complete *sty* (-rene degrading) operon consists of six genes (Figure 1.5 B): *stySc*, the styrene sensor component, *styR*, the transcriptional regulator, *styA* and *styB*, together catalyzing styrene epoxidation, *styC*, the styrene oxide isomerase, and *styD*, the dehydrogenase. The key enzyme for production of styrene oxide is the styrene monooxygenase *styAB* consisting of a larger oxygenase (StyA) and a smaller reductase (StyB) providing reduced FAD as a cofactor for the reaction (Otto *et al.*, 2004). StyAB could be cloned and heterologously expressed in other *Pseudomonas* strains such as KT2440 (Panke *et al.*, 1999) or in *E. coli* (Panke *et al.*, 2000; Park *et al.*, 2006) and yielded almost entirely (S)-styrene oxide with enantioselectivity of 99 %.

Styrene epoxidation was chosen as a model reaction because it is of high commercial interest, it involves hydrophobic reactants and products especially suitable for *Pseudomonas*, and it can serve as an example for many open questions regarding process feasibility and stability. However, the major interest of this work was to characterize population heterogeneity and its effect on bioprocesses. To this end, a C-terminal fusion of the *styA* gene with the fluorescent reporter Enhanced Green Fluorescent Protein (EGFP) was supplied by a cooperation partner within the project (Figure 1.5 C). It allows to track gene expression in living cells without the need to lyse cells and extract proteins biochemically. EGFP is one of hundreds of different fluorescent proteins known to date and is an improved variant (Cormack *et al.*, 1996) of the archetypal monomeric GFP isolated from *Aequoria victoria* and cloned in 1992 (Shimomura *et al.*, 1962; Prasher *et al.*, 1992).

1.5 Sources of heterogeneity in bioprocesses

Nearly all biotechnologically relevant microorganism propagate in a non-sexual 'clonal' manner. This means, a dividing mother cell will bring forth two genetically identical daughter cells. For reasons of simplicity, clonal populations have historically been regarded as homogeneous with its individuals being identical or at least highly similar. As a consequence, all morphologic and physiologic characteristics were supposed to be normally distributed within a population. However, countless exceptions from this rule have been found already, demonstrating that clonal cultures are by no means homogeneous but that heterogeneity is in fact the norm. Different types and characteristics of heterogeneity are outlined below, however, it should be kept in mind that some of the discussed concepts are overlapping with others, depending on the applied definition. For example, asymmetric cell division (section 1.5.2) can be a driver of heterogeneity on its own when there are dedicated cell types involved, but it may as well result in unequal distribution of cell constituents, a topic of its own (section 1.5.3).

1.5.1 Cell cycling

The most apparent cause of cell-to-cell heterogeneity is the need for cells to divide. To reach the required size to enter cell division, a mother cell must roughly duplicate its volume, DNA content and other cell constituents. A clonal population will therefore consist of a mixture of cells in different cell cycle stages, each showing a different morphology in terms of size and shape.

The knowledge of the complexity of cell cycle regulation in bacteria has grown in the last decades. Cell cycling is not a constantly running program, but cells adapt to variable environmental conditions by changing the pace or mode of cell cycling. In principle, the bacterial cell cycle
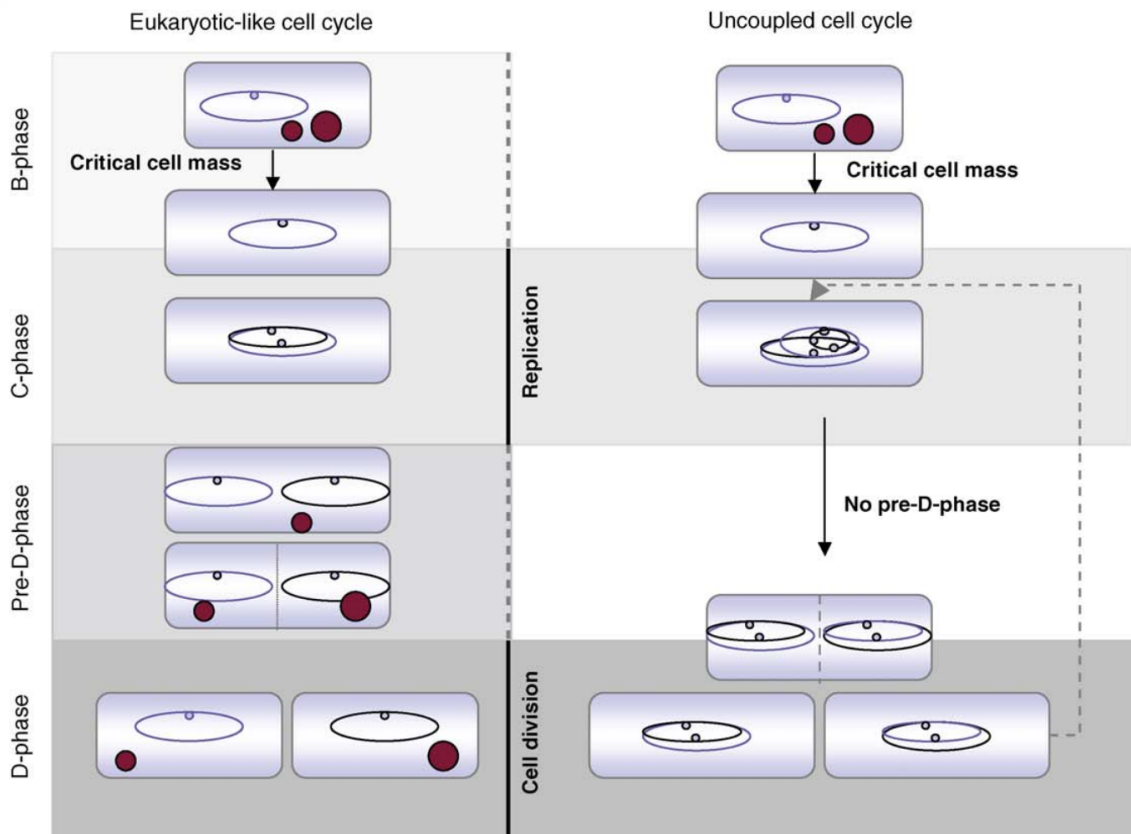
Figure 1.6: The bacterial cell cycle is regulated depending on nutrient availability. Cells with limited amounts of nutrients undergo eukaryotic-like cell cycling, with a constant duration of chromosome replication (C period) and cell division (D period), interrupted by periods of cell mass accumulation (B, left panel). In contrast, bacteria with unlimited nutrient supply will undergo uncoupled cell cycling, where new turns of DNA replication are initiated before previous ones are finished, resulting in larger cells with multiple chromosome content (right panel). Red circles: Storage compounds accumulated at different cell cycle stages. Figure reproduced from Müller *et al.* (2010).

can be split into three phases (Figure 1.6), the variable B period between birth and initiation of replication, the C period with replication of the chromosome, and the division period D (Wang & Levin, 2009; Müller *et al.*, 2010). If cell cycling progresses in an ordered fashion and replication rate is constant as defined by the fidelity of the DNA polymerase, how is it possible that cells with more than twice the amount of DNA were found? The answer was given by Cooper & Helmstetter (1968) who introduced the concept of multifork replication where new rounds of DNA replication are already initiated while previous rounds are not terminated yet. By this means, cells are able to adapt to high availability of nutrients, while the actual DNA replication rate (C period) remains constant. However, the B period disappears completely under maximum growth meaning that cells are constantly initiating new rounds of DNA replication. The division period D is constant again, limiting e.g. *E. coli* to a minimum generation time of 20 min. This uncoupled cell cycling is distinct from eukaryotic-like cell cycling, where replication must be

finished before cell division (Müller *et al.*, 2010). The model proposed by Cooper & Helmstetter (1968) was essentially confirmed and extended in later studies. It was found that C and D periods in *E. coli* are constant for short generation times but increase with generation times longer than 1 h (Skarstad *et al.*, 1985). In fact, microbial cells are able to control the initiation and speed of cell cycling by sensing the supply of nutrients and their own growth rate. In *E. coli*, the starting point for cell cycling is the initiation of DNA replication, which is directly regulated by the *oriC* binding protein DnaA, whose intracellular amount was shown to depend on growth rate and amino acid availability (Wang & Levin, 2009). Another supposed regulatory check point is the acquisition of cell mass. It was originally expected that DNA replication starts when cell mass has reached a certain 'critical' size, however, it was shown that this is true for some bacteria, but it is surely not a universal observation (Wang & Levin, 2009; Campos *et al.*, 2014).

1.5.2 Asymmetric cell division

Asymmetric cell division is a cell cycle related phenomenon which deserves attention in regard of population heterogeneity. In principle, asymmetric cell division is a tightly regulated process and must therefore be distinguished from unequal partitioning of cell constituents during cell division, which is a non-regulated 'random' process. Asymmetric cell division produces offspring that is morphologically and functionally different as an evolutionary strategy of specialization. The most prominent example employing this strategy is *Caulobacter crescentus*, leaving always one swarmer cell and one stalked cell after division (Figure 1.7 A). The swarmer cell will later become a stalked cell to undergo cell division itself. Five transcription factors globally regulating this unique cell differentiation are known (Quiñones-Valles *et al.*, 2014), however, the initial event sparking this regulatory cascade is unknown. Experimental evidence points towards unequal distribution of phosphate ions and proteolytic enzymes as a main actor (Tsokos & Laub, 2012).

Another example for asymmetric cell division is sporulation in the gram-positive bacterium *Bacillus subtilis* (Figure 1.7 B). Under limited nutrient conditions, a *B. subtilis* mother cell establishes a polar septum dividing the cell in two very different daughters, one large cell in regular state and one small spore of very different morphology and fate (Tan & Ramamurthi, 2014). The spore will eventually dehydrate and become dormant, being highly resistant to biochemical and physical stress (lysozyme, heat, UV irradiation). In the light of population heterogeneity, the onset of spore formation is highly interesting. Again, a global regulator (σF) is the key player for executing a differential gene expression program in the two future daughter cells (Tan & Ramamurthi, 2014). However, it is itself triggered by other proteins that act at polar septation of the cell. What ultimately causes the originally mid-cellular septum to polarize is unknown (Eswaramoorthy *et al.*, 2014).
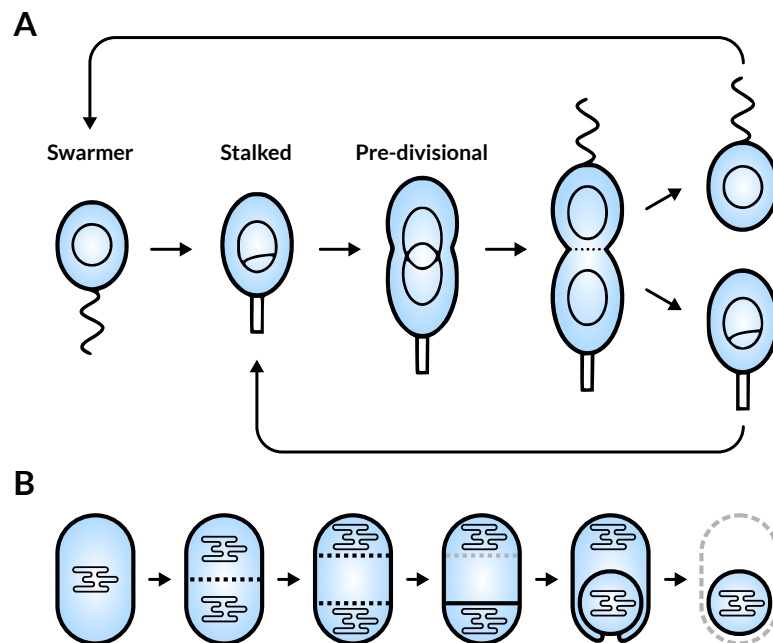
Figure 1.7: A) Cell cycle of *C. crescentus*, a bacterium producing two functionally different daughter cells after asymmetric division, a swarmer and a stalked cell type. Figure reproduced from Tsokos & Laub (2012). B) Spore formation in *B. subtilis* as another example of asymmetric division. A polar septum divides the mother cell into a large regular daughter cell and a small forespore. The mother cell is eventually lysed and the stress-resistent spore released. Figure reproduced from Tan & Ramamurthi (2014).

1.5.3 Unequal distribution of cell constituents

Unequal partitioning of cell constituents is a less obvious but presumably widespread phenomenon in the microbial world. It is caused by imperfect distribution of biomolecules during cell division and, contrasting to asymmetric cell division, is –with some exceptions– a random, non-regulated process. Two major mechanisms promote unequal distribution. The first one is partitioning noise, very similar to the gene expression noise concept, and it is due to low abundance of the respective molecule within the cell (Huh & Paulsson, 2011b). One can think of a regulatory protein being present in the mother cell with only five copies (Figure 1.8). Cell division will most likely distribute two and three copies to the daughter cells, respectively, but in some cases the difference will be even more extreme (1:4 or 0:5); such discrete distributions can be modeled using Poisson statistics. Interestingly, the heterogeneity detected in many studies dealing with noisy gene expression may actually be due to unequal partitioning (Huh & Paulsson, 2011a).

The second mechanism is biased distribution of molecules which are somehow spatially organized within the cell, and can affect also largely abundant molecules (Huh & Paulsson, 2011b). Examples are granules, membrane or nucleoid bound proteins, whose segregation must be ac-
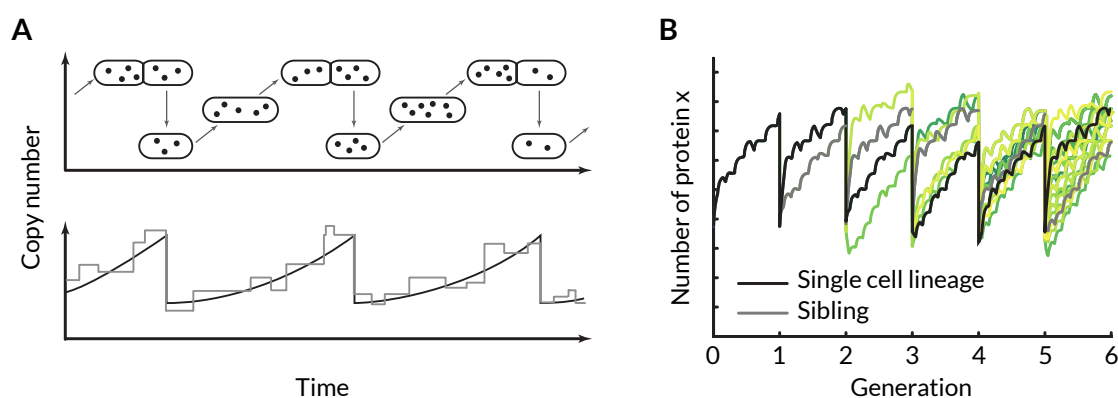
Figure 1.8: A) Cartoon of unequal partitioning of cell constituents, and model of the copy number per cell as average (black) and stochastic progression (grey). B) Following multiple cell divisions, unequal partitioning is virtually indistinguishable from gene expression noise. Figure reproduced from Huh & Paulsson (2011a).

tively controlled by the cells to avoid fatal errors (Henry & Crosson, 2013). Unequal distribution of molecules is also a factor discussed in aging of bacteria. In theory, perfectly even cell division would result in uniform immortal lineages of cells, however, it was proposed that 'old' and 'new' molecules are unevenly distributed between the maternal cell poles, eventually yielding one rejuvenated daughter cell and one aged ('maternal') daughter cell (Ackermann *et al.*, 2007). In fact, a *visually* apparently symmetric division in *E. coli* was shown to be *functionally* asymmetric, with maternal cells having a new and an old cell pole, and the offspring obtaining the old pole having a lower growth rate and higher incidence of mortality (Stewart *et al.*, 2005).

1.5.4 Gene expression noise

Gene expression in bacteria is a stochastic process. The mRNA and protein levels per cell for a particular gene can be very different over time or among different individuals in a population. The reason for that is *intrinsic* noise, the irregular initiation of transcription and translation, processes often involving a single locus of DNA and random binding of low-abundant regulatory proteins. A screening of single protein abundance for almost all genes of *E. coli* showed that most proteins appear in a range of 1 to 100 copies per cell (Taniguchi *et al.*, 2010), the lower range being prone to stochasticity. Gene expression noise was indeed high for proteins with a mean copy number less than 1, but declined to a minimum level for higher-abundant proteins. This minimum level of stochasticity was considered a 'background' signal, and was attributed to *extrinsic* noise, a term that includes all fluctuations of a dynamic physical and biological environment (Hilfinger & Paulsson, 2011). These impact of these fluctuations can vary from one cell to another, but is considered global for a single cell (Elowitz *et al.*, 2002).
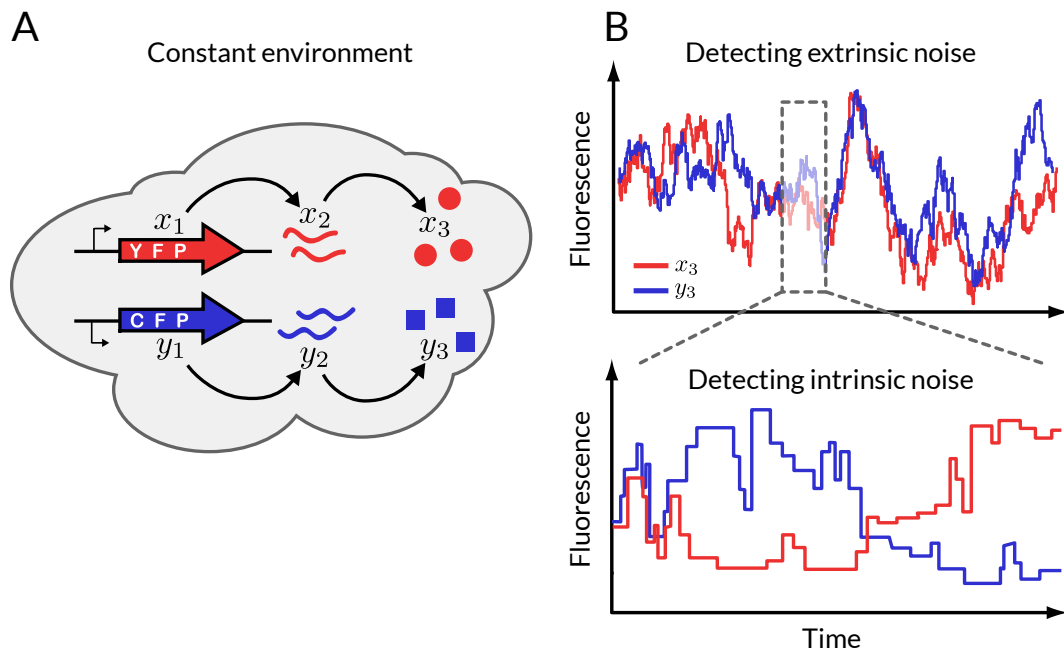
Figure 1.9: A) A dual reporter setup to detect intrinsic noise. B) Exemplary output of a dual reporter setup to differentiate intrinsic and extrinsic noise. x^{1-3}, y^{1-3}, gene, transcript and protein copy number for yellow (YFP) and cyan (CFP) fluorescent reporters, respectively. Figure modified from Hilfinger & Paulsson (2011).

Intrinsic (or gene expression) noise can be determined in principle by subtracting extrinsic noise using a dual reporter setup (Figure 1.9), first performed by Elowitz *et al.* (2002). It is based on the assumption that two identical gene expression systems with different fluorescent reporters should behave identical in a static environment, so that all variation in gene expression over time can be attributed to intrinsic noise. By this means, an equal contribution of intrinsic and extrinsic noise for a standard *lac*-promoted expression was shown, and the prediction of higher intrinsic noise for lower levels of gene expression was confirmed. However, experimental issues like differential maturation of the two fluorophores and DNA replication of the two reporter loci at different time points limit the significance of the dual reporter approach (Hilfinger & Paulsson, 2011). What is the biological meaning of stochastic gene expression? It might be a mere physical property in some respect, but it might as well be exploited by cells as a strategy to diversify, often called 'bet hedging'. The limited and discrete concentration of regulator proteins can, in conjunction with a feed forward circuit, lead to activation of pathways in *some* cells of a population but not *all* (Veening *et al.*, 2008). As an example, spore formation in *B. subtilis* is activated only in a subpopulation of early adopter cells sensing a worsening environment (Tan & Ramamurthi, 2014).

1.5.5 Mutations and chromosomal rearrangements

Heterogeneity can be caused by changes in the genetic information yielding a clonal, yet not genetically identical bacterial population. Many different sources of genetic heterogeneity are known, ranging from small scale point mutations to large scale chromosomal rearrangements. Point mutations are single base pair changes caused by erroneous DNA replication. Depending on the substituted base the mutation will be silent or alter the amino acid sequence of the protein. More dramatically, insertion or deletion ('indels') of one or more bases will probably destroy the reading frame of the protein. Mutations are rare events with rates of 10^{-6} to 10^{-10} per base pair and generation in bacteria (Rando & Verstrepen, 2007; Lee *et al.*, 2012; Wielgoss *et al.*, 2013). Recent studies nevertheless suggest that bacteria are able to tune mutation rates actively and therefore bypass Darwinian evolution (Rando & Verstrepen, 2007). It was found that the error prone SOS DNA replication in *E. coli* is specifically activated under deleterious conditions to increase genetic diversity (Petrosino *et al.*, 2009). Moreover, different *E. coli* strains (B and K12) acquired different mutations but in the very same genes to circumvent a metabolic bottleneck (Kim *et al.*, 2014). And finally, even in the absence of selection pressure mutation is biased in favoring strongly transcribed genes or one strand (direction of transcription) over the other (Juurik *et al.*, 2012).
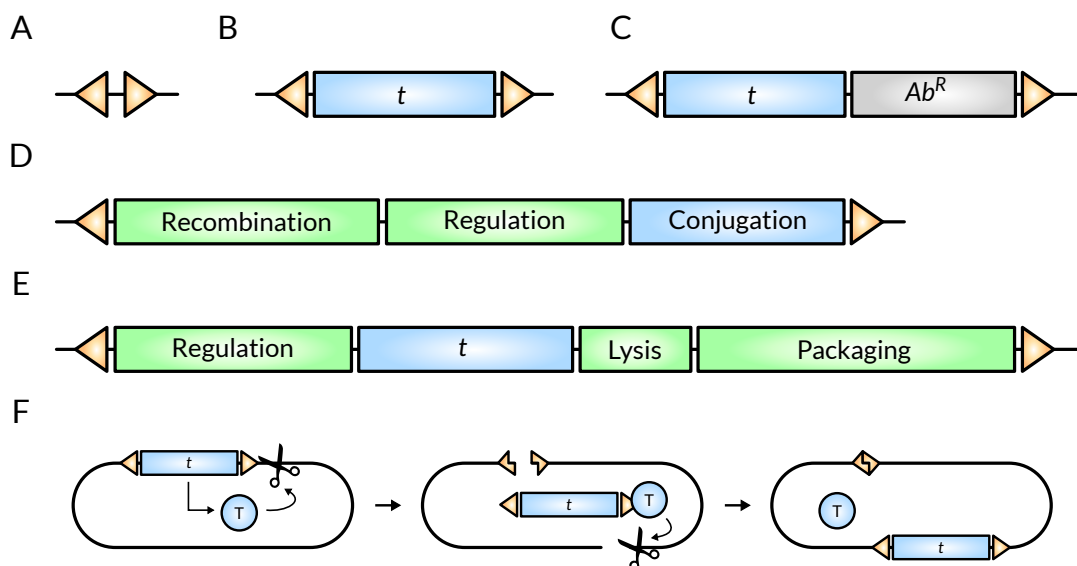
Figure 1.10: Mobile genetic elements are the main source of chromosomal rearrangements in bacteria. A) Miniature inverted repeat element (MITE) consisting of inverted repeats (triangles). B) Insertion (IS) element with additional transposase gene *t*. C) Transposon with additional antibiotic resistance genes (*Ab^R*). D) Integrative conjugative element (ICE), able to excise and transfer itself via conjugation. E) Lysogenic phage integrated into host genome and able to distribute itself via host lysis. F) Scheme of IS element jumping to another chromosomal locus leaving behind a 'footprint'. T, transposase. Figure modified from Darmon & Leach (2014).

On a larger scale, chromosomal rearrangements can drive evolution in a more dramatic, stepwise fashion. These rearrangements can be caused by intracellular processes such as homologous recombination or transposition, eventually leading to deletion, duplication, inversion or relocation of large chunks of DNA (Darmon & Leach, 2014). A vast number of genomic elements involved in DNA rearrangements are known, most of them comprising at least direct or inverted repeats as recognition sequences for recombinogenic enzymes (Figure 1.10). The space between recognition sites can be filled with functional genes determining the type of element, such as a plain transposase for insertion (IS) elements, additional host genes for (composite) transposons or nothing at all in case of miniature inverted repeat elements (MITEs) (Darmon & Leach, 2014). Very large and elaborate transposable elements are integrative conjugative elements (ICE) carrying not only genes for their own excision but also for conjugation and integration into other cells in the neighborhood (Reinhard & van der Meer, 2014). The same is true for lysogenic bacteriophages like μ with their complex lifestyle changing between a free extracellular and a host integrated state (Ranquet *et al.*, 2005). In a 25 years running experiment, Raeside *et al.* (2014) characterized the chromosomal rearrangements in 12 indepentent *E. coli* lineages grown for 40,000 generations. They found 110 events, mostly deletions (82) up to 55 kbp, but also 19 inversions comprising up to one third of the genome. Almost all rearrangements were due to the action of mobile elements.

1.5.6 DNA copy number fluctuation

A common source of bacterial heterogeneity is fluctuation of DNA copy numbers. Microorganisms regularly have variable numbers of chromosomes depending on the cell cycle stage (see 1.5.1) but also of extra-chromosomal DNA such as plasmids. Plasmids have become the major means of quick and powerful gene expression in research and industry as they are easy to modify, to transfer to a host cell, and can reach high copy numbers (Friehs, 2004). However, the plasmid copy number (PCN) can contribute significantly to population heterogeneity, as it is never uniform in a population but shows a certain variability from cell to cell or in individual cells over time (Wong Ng *et al.*, 2010). One reason is that the partitioning of plasmids during cell division is –with some exceptions– a stochastic process just like the partitioning of all other molecules (see 1.5.3). Therefore, modeling predicts a higher variability for smaller mean PCN, accompanied by a higher plasmid loss rate after partitioning (Figure 1.11, Summers (1991)). These assumptions were generally confirmed when comparing low copy to high copy replication systems (Wong Ng *et al.*, 2010), or plasmids equipped with a library of mutated replication systems displaying a PCN from 1 to 250 (Kittleson *et al.*, 2011).

One economically important aspect of PCN fluctuation is the appearance of plasmid free cells. Such cells will still divide, propagate, and consume resources without being productive, under-
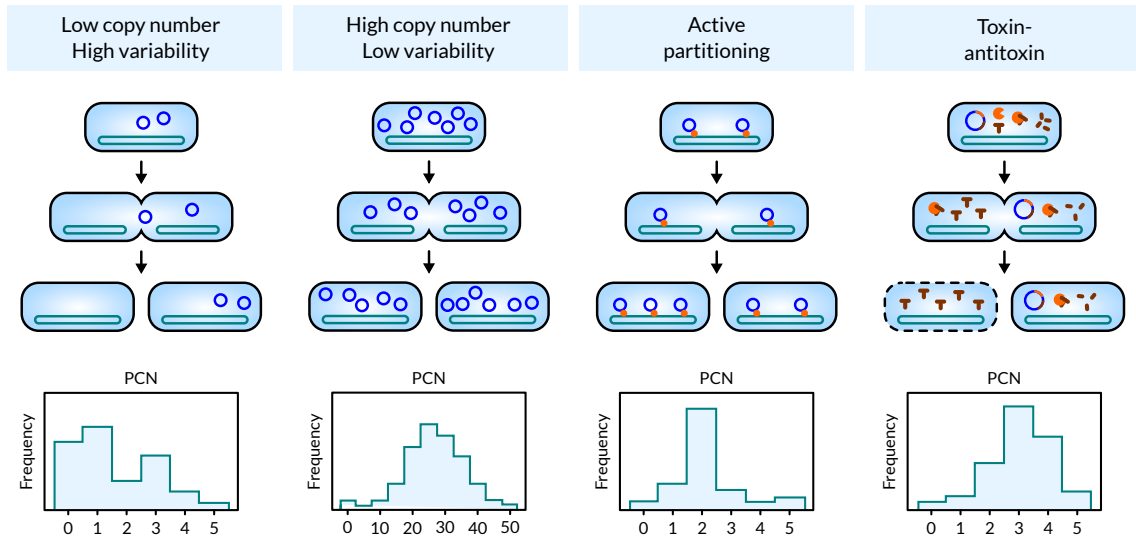
Figure 1.11: Population heterogeneity caused by plasmid copy number (PCN) fluctuation. Low copy plasmids show higher variability and rate of plasmid loss compared to high copy plasmids. Plasmids can be stabilized in a population by active partitioning or plasmid addiction systems. Histograms show exemplary copy number distributions.

mining a biotechnological process. In an evolutionary approach, two strains of *E. coli* carrying either a beneficial plasmid or nothing were mixed at different ratios and co-cultivated (Morris *et al.*, 2014). It was shown that only 20 % of plasmid-carrying cells remain after 75 generations, regardless of the mixing ratio; And without selection pressure, the plasmid-carrying subpopulation disappeared completely. This is usually attributed to plasmid loss during cell division, but sophisticated measurement of plasmid loss rates showed that the quick displacement of plasmid-carrying cells was in fact due to the growth advantage of plasmid-free cells, which is an accumulative effect while plasmid loss is only a linear function (Lau *et al.*, 2013). However, there are molecular mechanisms known to stabilize plasmid segregation (1.11). One is active partitioning, where (low-copy) plasmids are bound to the chromosome and hitchhike during cell division, thereby being equally partitioned to the two daughter cells (Lenz & Søgaard-Andersen, 2011). The other mechanism are plasmid addiction systems, where plasmids carry a factor on which the host cell desperately depends on. The best example are toxin-antitoxin systems: A plasmid carries one gene producing a toxin with long half-life and another gene for the antitoxin with short half-life. Upon plasmid loss, the antitoxin is degraded rapidly while the toxin (e.g. a porin or protease) unfolds its devastating action (Kroll *et al.*, 2010).

1.6 Methods to investigate heterogeneity

For decades microbial populations were characterized by using bulk measurements. This includes RNA or protein isolation for measuring gene expression, kinetic assays for productivity, or plate counting for viability. But to characterize heterogeneity, one requires assays to monitor cellular parameters on the single cell level. Already established techniques like (epi-) fluorescence microscopy or flow cytometry can be modified and extended to analyze single cells in evermore details and with higher throughput. In this work, flow cytometry was the fundamental technique to identify subpopulations of interest (1.6.2), while further techniques, subpopulation proteomics (1.6.3) and digital PCR (1.6.4), were adapted to be used on top for analyzing sorted cells.

1.6.1 Epifluorescence microscopy

The historically most important tool for single-cell analysis is fluorescence microscopy. A major breakthrough was the cloning of the Green Fluorescent Protein and all its successors, to enable a quantitative view on gene expression, protein localization, protein interaction and numerous other applications (Chudakov *et al.*, 2010). For monitoring heterogeneity, however, another equally important advancement was the advent of microfluidics to track cells over time without destroying the sample. An unimaginable diversity of microfluidic devices was already presented and therefore only a tiny selection of landmark experiments can be discussed here. For instance, a microfluidic device with 96 channels was used to probe gene expression noise in a library of 1,018 YFP tagged *E. coli* strains (Taniguchi *et al.*, 2010). This study was a significant contribution to pinpoint protein number and distribution in single cells. Microfluidics and fluorescence microscopy are a literally ideal combination to study cell lineages over long time. Wakamoto *et al.* (2013) could show that 'persisters', cells that survive antibiotic treatments in biofilms, are not dormant but actually divide and die with much longer cell cycle duration. Moreover, siblings are more likely to share the fate of survival or death, indicating an epigenetic, non-regulatory adaptation. Cells are usually entrapped in silica-based micro-chambers and adhere to the walls by direct physical contact. A novel contact-free single cell cultivation system is the 'envirostat', where cells are trapped in an electromagnetic field (Fritzsch *et al.*, 2013). Using this device, it could be shown that different microbial species are able to grow significantly faster than in conventional space-limited cultivation systems (Dusny *et al.*, 2012).

1.6.2 Flow cytometry and cell sorting

Flow cytometry is an established technique for the analysis of single cells. It is based on the optical measurement of cells entrapped in a narrow fluid stream, separately passing one or more laser
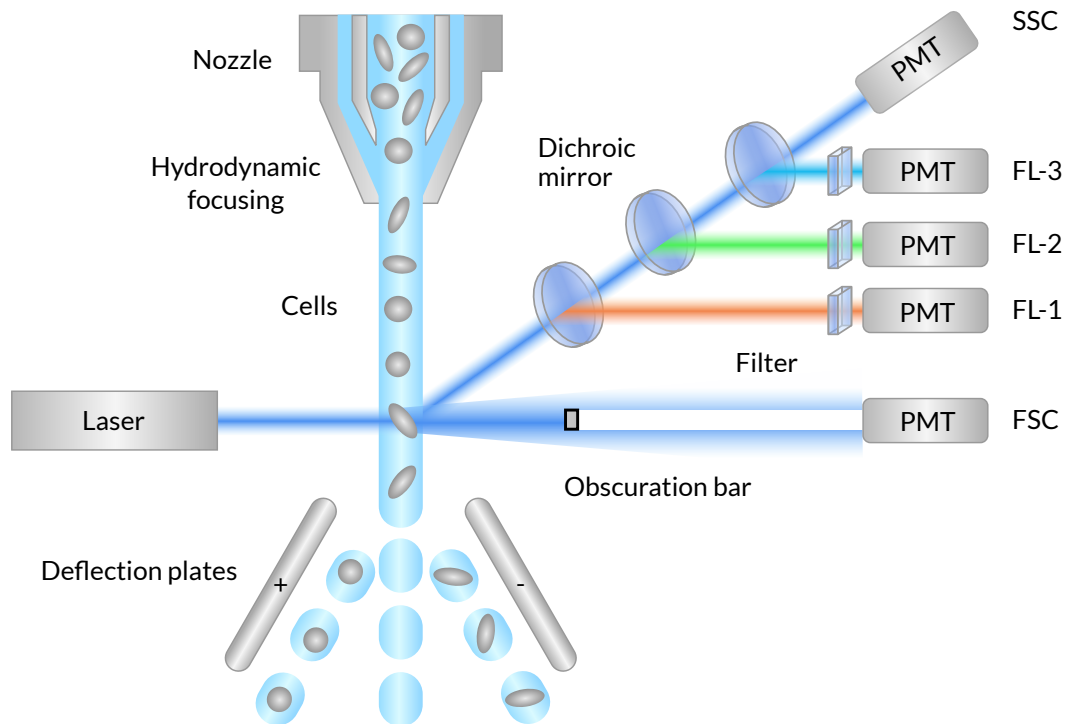
Figure 1.12: Scheme of optical system of a cytometer. Cells are focused in a narrow stream of sheath fluid and pass a laser beam. Fluorescence signals (FL) of different wavelength are detected in photomultipliers (PMT) by splitting the emitted light by dichroic mirrors. Cell sorting is achieved by converting the laminar sample stream into single droplets, which can be individually deflected by an electromagnetic field.

beams. The light scattering in the forward and sideward direction (FSC and SSC, respectively) can be detected as well as fluorescence emission at arbitrary wavelengths (Figure 1.12). Highly multi-parametric detection of fluorescent antibodies is nowadays a standard procedure in medical research (Chattopadhyay & Roederer, 2012), however, this is not directly applicable to bacteria because of their rigid cell wall and the lack of specific antibodies (Müller & Nebe-von Caron, 2010). Flow cytometry has been extensively used in biotechnological settings to screen for productivity using properties such as light scattering, fluorescent reporter genes or intracellular RNA aptamer probes (Hedhammar *et al.*, 2005; Strovas & Lidstrom, 2009; Wyre & Overton, 2014; Siedler *et al.*, 2014). It has been used as well to study cell cycling or viability via staining of DNA or membrane integrity, respectively (Skarstad *et al.*, 1985; Hewitt *et al.*, 2007; Want *et al.*, 2011). In comparison to microscopic single cell analysis, some disadvantages of flow cytometry come to mind: A repeated measurement of the same cells is not possible as cells are not recycled after analysis, no detailed images are acquired but only single values per parameter, and in consequence the intracellular localization of fluorophores remains unknown. However, the strength of flow cytometry lies elsewhere. First, extremely high throughput enables analysis of up to 100,000 cells per second, which allows quantitative analysis of even the rarest events

over several orders of magnitude. Second, flow cytometric analysis can be complemented by cell sorting. This allows retrieval of intact living cells of a subpopulation of choice with high purity. Cell sorting can be based on deflection of droplets containing the particle of interest, or based on mechanically diverting a fluid stream, which is suitable for larger particles like plant seeds or *Drosophila* eggs (Shapiro, 2003). More recently, microfluidics were as well set up for cell sorting by using e.g. acoustic, electric or magnetic fields, although at a lower maximal speed compared to conventional sorting (Shields *et al.*, 2015).

1.6.3 Proteomics using mass spectrometry

Analyzing heterogeneity by optical means is limited to morphological characteristics of cells while staining of nucleic acids or membrane integrity allows to retrieve additional physiological data. However, to get a more detailed picture of the underlying molecular basis for heterogeneity, flow cytometry can be used as a first step to identify and sort subpopulations of interest. In a second step the sorted cells are subjected to another type of analysis, here the simultaneous detection of many cellular proteins by liquid chromatography tandem mass spectrometry (LC-MS-MS). Mass spectrometry is a widely used technique to map the proteome of microbial cells and was already used for *Pseudomonas* strains (Nikodinovic-Runic *et al.*, 2009; Wijte *et al.*, 2011; Yun *et al.*, 2011). Of the 5350 ORFs predicted *in silico* in *P. putida* KT2440, roughly 25 % (up to 1286 proteins, Yun *et al.* (2011)) can be found expressed as proteins. However, subpopulations of microbial cells obtained by sorting have rarely been used for mass spectrometry. In a proof-of-principle experiment, it was possible to separate *E. coli* and *P. putida* from an artificial mixture for further LC-MS-MS analysis (Jehmlich *et al.*, 2010). Another application was the extraction of the intracellular pathogen *Staphylococcus aureaus* from lysed host cells by sorting prior to LC-MS-MS analysis (Surmann *et al.*, 2014). However, sorting morphologically diverse subpopulations of one and the same microbial species for subsequent mass spectrometry was not performed before, and efforts were therefore undertaken in this study to establish and refine this method (Aims in section 1.7).

1.6.4 Digital PCR to quantify DNA copy number

During the course of this work it became obvious that differences in productivity of a *Pseudomonas* expression strain, as determined by EGFP fluorescence, were linked to plasmid copy number fluctuation. The fluorescence intensity of a reporter protein such as EGFP is unfortunately not directly correlated to the number of gene copies or transcripts within the cell. Therefore, a second technique ought to be established to monitor even small differences in plasmid copy number of template cells obtained by cell sorting: Digital polymerase chain reaction or digital PCR. It is a novel technique based on the partitioning of a standard PCR reaction volume into many small

compartments, in each happening the PCR reaction individually (Vogelstein & Kinzler, 1999). Using the most advanced system, droplet digital PCR (ddPCR), the result is a number of negative droplets which contain no template, and a number of positive droplets containing one or more template molecules (Hindson *et al.*, 2011). The positive droplets are simply counted and the *absolute* number of template molecules can be calculated using Poisson statistics. This simple probabilistic model takes the occupancy of droplets by more than one template into account. This work flow is extremely straightforward compared to the *relative* quantification by conventional quantitative real time PCR (qRT-PCR), where the result is a fluorescence curve which is compared to a known standard (Whale *et al.*, 2012). Digital PCR has proven to be more accurate than qRT-PCR and less prone to inhibitory substances (Strain *et al.*, 2013; Hindson *et al.*, 2013). However, it has not been used in conjunction with cell sorting, although it is literally predestined for an application providing only little amounts of template.

1.7 Aims of this study

To establish techniques for the single cell analysis of *Pseudomonas putida* in biotechnological settings. The model strain used throughout this work was *P. putida* KT2440 transformed with different plasmid vectors for controlled expression of the StyA-EGFP fusion protein and StyB. Specific aims were the flow cytometric analysis of light scatter properties, DNA content using DAPI staining, and EGFP fluorescence to characterize the physiological state and productivity of the cells.

To investigate microbial heterogeneity in a model bioprocess applying subpopulation proteomics. First, different methods for fixation and storage of intact *Pseudomomas* cells were tested in order to preserve cell morphology and protein composition for flow cytometric sorting and mass spectrometry, respectively (publication 1). Second, the use of subpopulation proteomics was intended to elucidate the cause for a striking heterogeneity in StyA-EGFP producing populations (publication 2). And third, it was intended to quantify the molecular basis of heterogeneity in cells undergoing cell cycling. To this end, subpopulations with different DNA content appearing under strictly defined growth conditions during chemostat cultivation were sorted and analyzed via mass spectrometry (publication 3).

To reveal the molecular cause for heterogeneity in StyA-EGFP expression. The previous proteomic study clearly pointed towards differences in plasmid copy number as a source for population heterogeneity. Therefore, a new method should be established for precise quantification of DNA *via* digital PCR on the subpopulation level. Second, this technique should be used to test the hypothesis of plasmid loss in a subpopulation of non-productive cells (publication 4).

2 Publications

2.1 Overview of publications

Publication 1

Jahn, Michael; Seifert, Jana; Hübschmann, Thomas; von Bergen, Martin; Harms, Hauke; Müller, Susann. *Comparison of preservation methods for bacterial cells in cytomics and proteomics*. Journal Of Integrated Omics 3:1-9, **2013**.

Significance: In this study three different methods for whole cell preservation were tested, azide fixation, deep freezing and vacuum drying. All three methods were already known, however, they have not been explicitly tested with cell sorting or protein mass spectrometry in mind. It was shown that deep freezing preserved the cells best for at least one month, while vacuum drying performed almost equally well. However, azide fixation was regarded as deleterious for protein composition. The results of this study were an essential prerequisite for the following experiments.

Publication 2

Jahn, Michael; Seifert, Jana; von Bergen, Martin; Schmid, Andreas; Bühler, Bruno; Müller, Susann. *Subpopulation-proteomics in prokaryotic populations*. Current Opinion in Biotechnology 24:79-87, **2013**.

Significance: In the first part this review discusses methods for single cell proteomics, concluding that only limited information about protein composition of a single cell can be obtained. In the second part, a case study is presented using *P. putida* KT2440 as a model organism showing discrete subpopulations during heterologous gene expression. For the first time, subpopulations of a bacterium are sorted and analyzed by mass spectrometry, yielding quantitative data of almost 800 proteins.

Publication 3

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

Significance: This study follows the lead of the previous publication. Applying mass spectrometry to subpopulations of different DNA content, it was shown for the first time that cells of the same population but in different cell cycle stage are essentially identical. However, strong differences were found between populations of different growth rate, demonstrating that this is a major parameter determining cell physiology.

Publication 4

Jahn, Michael; Vorpahl, Carsten; Türkowsky, Dominique; Lindmeyer, Martin; Bühler, Bruno;Harms, Hauke; Müller, Susann. *Accurate determination of plasmid copy number of flow-sorted cells using droplet digital PCR*. Analytical Chemistry 86:5969-76, **2014**.

Significance: The results of publication two pointed towards plasmid copy number as a driver of heterogeneity. This study for the first time demonstrates exact measurement of plasmid copy number in sorted subpopulations using novel droplet digital PCR. As little as 1,000 sorted cells were shown to provide the highest accuracy. Finally, the missing productivity of cells could indeed be attributed to plasmid loss.

2.2 Published articles

- Publication 1 can be found at page 34.

- Publication 2 can be found at page 43.

- Publication 3 can be found at page 52.

- Publication 4 can be found at page 62.
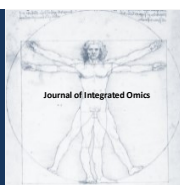
ORIGINAL ARTICLE

Comparison of preservation methods for bacterial cells in cytomics and proteomics

Michael Jahn[1], Jana Seifert[2], Thomas Hübschmann[1], Martin von Bergen[2,3,4], Hauke Harms[1], Susann Müller[1]

[1]Department of Environmental Microbiology, UFZ - Helmholtz Centre for Environmental Research, Permoserstr. 15, D-04318 Leipzig, Germany; [2]Department of Proteomics, UFZ - Helmholtz Centre for Environmental Research, Permoserstr. 15, D-04318 Leipzig, Germany; [3]Department of Metabolomics, UFZ - Helmholtz Centre for Environmental Research, Permoserstr. 15, D-04318 Leipzig, Germany; [4]Department of Biotechnology, Chemistry and Environmental Engineering, Aalborg University, Sohngaardsholmsvej 49, DK-9000 Aalborg, Denmark

Abstract

Cell sampling during long-term experiments usually requires reliable storage of cells for later analysis. In this study, we evaluated three different preservation strategies (sodium azide fixation, deep freezing and vacuum drying) with regard to their effects on bacterial cells. *Pseudomonas putida* was used as a model organism and stored for shorter (2 d) and longer periods (28 d). The impact of the treatments (preservation method, duration) was evaluated on the level of single cells using flow cytometry and on the population level using protein mass spectrometry. On the single cell level, the effect of sodium azide fixation was found to be small ($1.01 \leq sd \leq 1.76$) for short term and larger for long term storage ($1.59 \leq sd \leq 2.33$), as determined by FlowFP fingerprinting. In contrast, deep frozen and vacuum dried cells showed properties highly similar to fresh reference cells, even after extended storage ($0.5 \leq sd \leq 1.2$). On the population level, the mass spectrometric analysis revealed about 800 proteins for each sample and storage condition. The proteome profiles evaluated by label-free quantification showed that variation within functional groups was least for deep frozen and vacuum dried cell samples after 2 d (sd log2 relative protein quantity < 1) and marginally increased after 28 d. In contrast, sodium azide fixation caused higher variations between functional groups although the number of detected proteins and the respective peptide coverage excluded protein degradation. In conclusion, deep freezing was found to be the method of choice, but simple vacuum drying of cells with storage at 4°C can be a convenient alternative.

Keywords: Fixation; Cryopreservation; Vacuum drying; Bacteria; Flow cytometry; Mass spectrometry.

Abbreviations:

F, fresh; VD, vacuum drying; DF, deep freezing; SAF, sodium azide fixation; FCM, flow cytometry; MS, mass spectrometry

1. Introduction

The number of samples that can be analyzed by flow cytometry is steadily increasing and recently reached 1536-well plate format as a platform for parallel analysis [1]. The advent of high-throughput analyses raises the question of efficient sample preservation, both for plain flow cytometry and for cell sorting in combination with other *Omics* techniques in microbiology [2]. With such applications in mind, the recovery of whole intact cells after storage is essential for flow cytometric analysis and the reliability of this recovery is particularly important for long-term experiments. The process of sample preservation consists of two parts, sample preparation and sample storage. Here, we define sample preparation as a measure to stop cellular activity and prevent cells from alteration or decay. Usually relatively harsh methods derived from histology were used for preparation, commonly known as fixation. For instance,

*Corresponding author: Susann Müller, UFZ - Helmholtz Centre for Environmental Research, Permoserstr. 15, D-04318 Leipzig, Germany; Phone number: +49 341 235 1318; E-mail address: susann.mueller@ufz.de

alcohols and aldehydes have the advantage of inactivating biohazardous specimens [3]. In general, different fixation and preparation techniques have different benefits and drawbacks, as they change the sample in characteristic ways. The risk of further changes caused by subsequent storage until sample analysis is mainly determined by the storage time and temperature [4]. The diversity of preparation techniques and storage conditions is large, and the choice depends on the kind of analysis samples are subjected to.

The most widely-used fixation methods for microbial cells are alcohol- or aldehyde-based methods, which mainly work via protein cross-linking (aldehydes), cell permeabilization and water removal (alcohol) [3, 5]. Although aldehydes permit easy fixation and storage even at room temperature (RT), major drawbacks include alterations of cell morphology, cell aggregation, loss of biomolecules and increased autofluorescence [3, 6, 7]. Another chemical preservation method, which can be also regarded as fixation, is the use of (sodium-) azide, which was proven useful for microorganisms [8, 9]. Azide inhibits the terminal enzyme of the respiratory chain and thereby prevents energy production of aerobic cells [10-12]. Metal salts like barium, nickel and molybdenum in combination with 10% sodium azide proved especially useful for anaerobic bacteria and provided highly reproducible cytometric patterns for at least nine days [13, 14]. These chemical preservation techniques are usually sufficient for flow cytometry, but destructive effects like protein cross-linking in case of aldehydes disqualify a method for protein mass spectrometry. A method widely used in biobanking is cryopreservation by deep freezing (DF) at temperatures below -60°C with addition of a cryoprotective agent, such as glycerol or sugars [15-17]. Deep freezing was suitable for preservation of methane oxidizing bacteria as tested by flow cytometry [18], but performed poorly for natural microbial communities [13]. In the food and pharmaceutical industry, drying of bacterial cells by different techniques is used for product conservation [19, 20]. Freeze drying, spray drying or low temperature (~0°C) vacuum drying were shown to preserve different bacterial strains [21-23], but have not been applied in single cell analysis.

In this study, we compared three different preparation and storage procedures suitable for single cell analysis and simple enough for standard laboratory application: Sodium azide fixation (SAF) followed by storage at 4°C, vacuum drying (VD) followed by storage at 4°C, and deep freezing (DF) with liquid nitrogen and subsequent storage at -80°C. The bacterial strain *Pseudomonas putida* was used as model organism for analysis by flow cytometry (FCM) and by protein mass spectrometry (MS). The combination of both techniques is promising, as cells with different characteristics can be identified and sorted by FCM, and the subpopulation's proteome can be analyzed by MS. Subpopulation proteomics was already performed for an artificial mixture of bacteria [24] and for pure cultures of *P. putida* [2]. Besides MS, FCM can also be coupled with other

downstream applications like RNA isolation [25, 26]. In the presented study, FCM and MS were used to evaluate the effect of the selected preservation methods (VD, DF, SAF) on cell characteristics (DNA content, light scatter) and protein profile, respectively, in comparison to fresh samples (F). A gate-free similarity fingerprinting method was applied for analyzing cytometry data [27] and a novel functional clustering algorithm was used for comparison of proteome profiles.

2. Materials and Methods

2.1. *Bacterial strains and cultivation conditions*

Pseudomonas putida KT2440 (source: DSMZ – German Collection of Microorganisms and Cell Cultures) was cultivated overnight at 30°C on a rotary shaker (180 rpm) in minimal medium (Na_2HPO_4 6 g/L, KH_2PO_4 3 g/L, NaCl 0.5 g/L, NH_4Cl 1 g/L, $MgSO_4$ 0.5 g/L, $CaCl_2$ 15 mg/L, $ZnSO_4$ x $7H_2O$ 3.6 mg/L, $CuSO_4$ x $5H_2O$ 0.625 mg/L, H_3BO_3 0.15 mg/L, $FeSO_4$ x $7H_2O$ 6 mg/L, $CaCO_3$ 5 mg/L, $MnSO_4$ x $7H_2O$ 3 mg/L, $CoSO_4$ x $7H_2O$ 0.7 mg/L) with 2 g/L glucose as sole carbon and energy source. A 500-ml shaking flask with 100 ml of minimal medium was inoculated with a volume of overnight culture corresponding to an initial optical density of 0.05 at 600 nm (OD_{600nm}, $d_{cuvette}$ = 0.5 cm). For comparison, *Escherichia coli* DH5α (source: DSMZ) was cultivated overnight at 37°C on a rotary shaker (180 rpm) in LB medium. A 500 ml shaking flask with 100 ml of LB medium was inoculated to an initial OD_{600nm} of 0.05. The growth of the cells was monitored by measurement of OD_{600nm} up to a maximum incubation time of 12 h.

2.2. *Cell preparation and storage*

Cells of *P. putida* were collected every two hours for a total of 12 hours after inoculation covering all cell cycle stages (Figure 1B). For *E. coli*, cells from the lag and exponential growth phases at 0, 1 and 2 hours were used. Samples were taken at various time points by centrifugation of up to 2 ml cell suspension in a microcentrifuge (Heraeus Fresco 21) for 5 min at room temperature and 5,000 x g, and the supernatant was discarded. For DF, the cells were resuspended in 1 ml phosphate buffered saline (PBS; 6 mM Na_2HPO_4, 1.8 mM NaH_2PO_4, 145 mM NaCl, pH 7.2) containing 15 % (v/v) glycerol as cryoprotective agent, incubated for 10 min on ice and shock frozen in liquid nitrogen for subsequent storage at -80°C. For VD, any residual medium was removed and the cell pellet was vacuum dried for 30 min at 30°C using a vacuum concentrator (N-Biotek Micro-Cenvac). The dried cell pellet was stored at 4°C. For SAF, the cells were resuspended in 2 ml 10 % (v/v) sodium azide (NaN_3) and stored at 4°C.
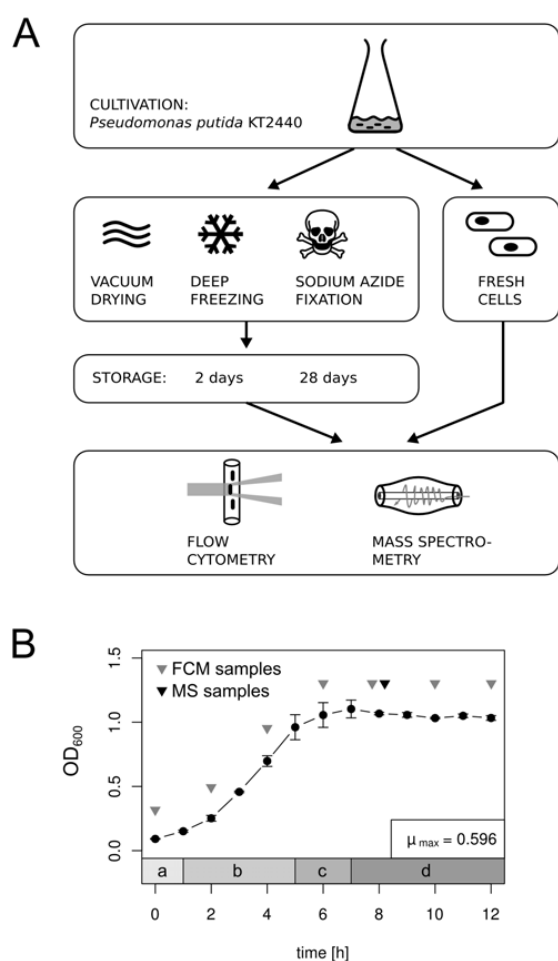
2.3. *Cell staining*

Figure 1. Experimental design of the study. (**A**) Bacterial cultures were grown in shaking flasks, samples were taken at different time points and prepared for storage by vacuum drying (VD), deep freezing (DF) and sodium azide fixation (SAF). The samples were stored for 2 d or 28 d and then analyzed by flow cytometry and mass spectrometry. Fresh samples (F) were directly analyzed and served as reference. (**B**) Growth of *P. putida* KT2440 was monitored by optical density measurement (OD_{600nm}, $d_{cuvette}$ = 0.5 cm) and covered lag phase (a), exponential (b), early stationary (c) and stationary (d) growth phase. Samples for flow cytometry were taken every 2 h (▼) and for mass spectrometry at 8 h of growth

Fresh (only *P. putida*), deep frozen and sodium azide fixed cells were centrifuged for 5 min at RT and 5,000 x g and the supernatant was discarded. The resulting cell pellets as well as VD pellets were resuspended in ice cold PBS by repeated pipetting, and the OD_{600nm} was adjusted to 0.05. For DNA staining, 1 ml of the cells was harvested by centrifugation, and the pellet was resuspended in 1 ml of permeabilization buffer (0.3 M citric acid, 5 g/L Tween 20) and incubated for 10 min on ice. After centrifugation the supernatant was removed and the cells were resuspended in 1 ml DNA staining solution (0.68 µM 4',6-diamidino-2'-phenylindole

(DAPI), 0.4 M Na_2HPO_4) [28] and incubated for at least 15 min at RT. Prior to flow cytometry, cell clusters were removed by filtration using a membrane of 50 µm pore diameter (CellTrics, Partec) to prevent clogging of the cytometer nozzle. A pumping of DAPI by fresh cells, as reported for several nucleic acid stains [29], was not observed.

2.4. *Flow cytometry and data analysis*

Flow cytometry was performed using a MoFlo cell sorter (Beckman-Coulter, USA) as described in [24]. The DAPI fluorescence was determined using a multi-line UV laser (333–365 nm, 100 mW) for excitation, and emission was detected in the FL4 channel (450±30 nm). Prior to all measurements, the instrument was adjusted using fluorescent beads and a biological standard. Data were recorded with the Summit v4.3 software (Beckman-Coulter) and further analyzed using the Bioconductor framework for R [30]. The electronic noise was removed (forward scatter, threshold of 25) to prevent bias of similarity analyses. For better comparability, the dominant peaks of the forward scatter (FSC) and DAPI (FL4) channels were normalized (warpSet function, variable peak number, grouping on parameter time) using the Bioconductor *flowStats* package [31]. Similarity fingerprinting with the FSC and FL4 channels (standard deviation method, five recursions) was performed using the *FlowFP* package [27].

2.5. *Identification of proteins by LC-MS-MS*

For proteomics, cells of *P. putida* KT2440 (three technical replicates) were harvested after 8 h of growth and either stored as described above or directly prepared for MS by resuspension in 2 ml PBS. The cell number was determined using flow cytometry and a sample volume corresponding to 1 x 10^8 cells was harvested by centrifugation and resuspended in 25 µl 25 mM ammonium bicarbonate buffer (NH_4HCO_3, pH 7.8) with 1 µl acetonitrile and 5 µl of trypsin (0.25 µg/µl, Promega, Madison, USA) for proteolytic digestion. The samples were incubated over night at 37°C with continuous shaking (180 rpm) and the digestion was stopped by addition of formic acid (FA, 0.1 % (v/v) final concentration). Cell debris was removed by centrifugation for 10 min at 13,000 x g and RT and the supernatant was transferred to a fresh 0.5 ml-tube. Samples were stored at -20°C until analysis. The peptide solution was then purified using the ZipTip protocol (Millipore, Bedford, USA), dried using a vacuum concentrator and the remaining peptides resuspended in FA. The solution was sonicated in a water bath for 5 min prior to injection. Peptides were separated and measured by a high-pressure liquid chromatography (nano-UPLC) system (nano-Acquity, Waters) coupled to an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) as described in [32]. Continuous scanning of eluted peptide ions was carried out

between 300-1600 m/z, automatically switching to MS/MS CID mode on ions exceeding an intensity of 3000.

The retrieved raw data were analyzed by MaxQuant (version 1.2.2.5) [33] with the genome sequence of *P. putida* KT2440 as the database. The MaxQuant settings can be found in more detail in supplementary Table S1. The label-free quantification (LFQ) values were used for protein quantification and can be found in supplementary Table S3. For peptide mapping, the original mass spectra were further analyzed by Thermo Discoverer (v.1.2.0.208), Mascot (v 2.3) and the NCBInr database (as of February 2013) with a restriction to sequence entries of *P. putida* KT2440 (available in supplementary Table S4).

2.6. Proteome mapping

The mean and standard deviation of the obtained LFQ values were calculated for each triplicate. The relative protein quantity was calculated as ratio of the protein quantity of the respective stored sample to the fresh sample (F). To add biological information to the detected proteins, their subcellular localization was predicted *in silico* by PSORTb v3.0 [34]. Furthermore, proteins were annotated using the KEGG BRITE functional hierarchy for *P. putida* KT2440 (http://www.genome.jp/kegg-bin/get_htext?ppu00001.keg, [35]), adding four hierarchical levels (here called 'system', 'process', 'pathway' and 'protein'). Sunburst treemaps were created using a custom recursive function in R (supplementary information S2).

3. Results

3.1. Cell preparation and storage

The influence of preservation method and storage time on the stability of bacterial cells was investigated. To this end, three different cell parameters (FSC, SSC and DNA content) were selected as indicators and analyzed by FCM. Likewise, the protein profile of the same samples was analyzed by MS. Samples of *P. putida* KT2440 were taken at various time points during batch cultivation. Hence, the cells display a different morphology, DNA content and possibly also sensitivity to preparation methods due to various growth stages. At every time point, cells were either directly analyzed by flow cytometry (fresh, F) or after preservation by VD, DF or SAF as depicted in Figure 1A. Since VD has not been used for cell preservation in cytometry before, the method was optimized regarding drying temperature and duration. Out of three different drying durations (10, 30, 60 min) and two temperatures (30°C, 60°C), the most distinct distribution (FSC, DNA) was obtained at 30°C for 30 min (supplementary Figure S1). Likewise, DF was tested with 15 and 50 % (v/v) glycerol in PBS as a cryoprotective agent, but no difference was observed (supplementary Figure S2).

3.2. Analysis of flow cytometric pattern similarity

Samples of *P. putida* KT2440 from seven time points were prepared by four different methods (F, VD, DF, SAF) and either directly analyzed (F) or stored for 2 d and 28 d (VD, DF, SAF). The light scattering and DNA content of the cells were analyzed by flow cytometry (Figure 2A). The side scatter signal showed no remarkable differences between samples and was disregarded for similarity analysis.
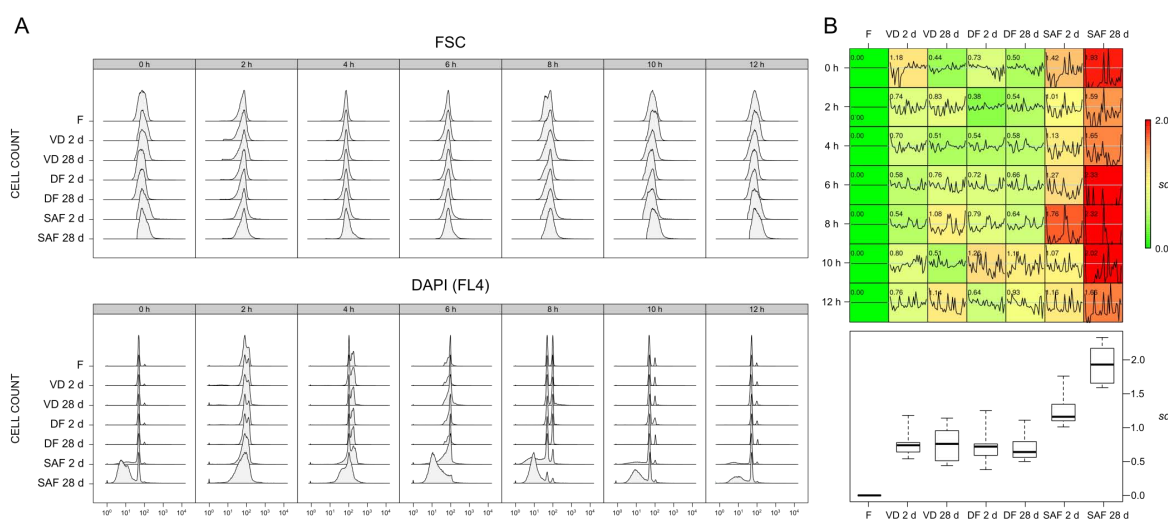
Figure 2. Similarity analysis of stored *P. putida* KT2440 cells using flow cytometry. The cells were either directly analyzed (F) or prepared by vacuum drying (VD), deep freezing (DF) and sodium azide fixation (SAF) and stored for 2 d or 28 d. (**A**) Histograms of forward scatter (FSC, upper panel) and DAPI fluorescence (FL4, lower panel) after 0-12 h of incubation. (**B**) Similarity fingerprint of stored samples in comparison to the fresh sample (F). Given is the standard deviation *sd* as index for bin differences, which increases with increasing dissimilarity. The retrieved *sd* values for each method are summarized in a box plot.

Predominantly, the FSC signal showed a unimodal distribution, which shifted according to the growth of the cells. The cellular DNA content changed during cultivation, ranging from a single chromosome equivalent (C1n) to two or more copies ($\geq$ C2n). The DNA pattern of the stored samples was highly similar to the fresh samples for VD and DF, but SAF treated cells showed a certain proportion of cells below the C1n peak after 2 d, which further increased after 28 d of storage. To quantify similarity of flow cytometric patterns a fingerprinting method (*FlowFP*) was employed, involving FSC and DNA (Figure 2B). It is based on a probability binning algorithm, which constructs a model from a reference sample based on the distribution of events in it. Here, the respective fresh sample (F) of each time point served as the reference sample. This is then compared to the stored samples yielding an index for similarity, the standard deviation of bin differences (*sd*). It is 0 for identical samples and increases with increasing dissimilarity. The sensitivity limit of the method was determined by comparison of two virtual halves of the same sample F, yielding an $sd \leq 0.25$ (supplementary Figure S3). An equal similarity to the fresh reference sample was found for VD and DF ($0.5 \leq sd \leq 1.2$, Figure 2B) independent of the storage duration. Samples stored by SAF were less similar after 2 d ($1.01 \leq sd \leq 1.76$) and particularly after prolonged storage of 28 d ($1.59 \leq sd \leq 2.33$). The growth phase of *P. putida* had only a small influence on the similarity to the reference sample. For comparison, a similar experiment was conducted using another bacterial strain, *E. coli* DH5α. In contrast to *P. putida*, cells of the exponential growth phase

were more amenable to alteration than lag phase cells (supplementary Figure S4). Furthermore, the deleterious effect of SAF was much less pronounced for *E. coli* than for *P. putida*.

3.3. Identification of proteins by MS

Flow cytometry is a powerful application for analysis and sorting of cells, and can be readily combined with other downstream applications such as proteomics. However, the storage of cells until sorting may influence their protein profile. Therefore, we tested the ability of the selected methods (DF, VD, SAF) to preserve the protein composition of *P. putida* cells for two time periods (2 d, 28 d). Cell samples were acquired at the early stationary growth phase (8 h) and either stored, or instantly analyzed serving as the reference (F). The cellular proteins were analyzed by shotgun mass spectrometry with label-free quantification. A total number of 971 different proteins was detected across all samples, with a range of 793 to 858 different proteins present in at least one replicate per single sample (Figure 3A). The number of proteins common to all samples was 751 (intersection, IS), indicating comparably good protein recovery for all preservation techniques, although the number of recovered proteins across replicates varied by each method (supplementary Figure S5). Particularly, samples stored for 28 d with VD and SAF showed a reduced number of detected proteins per replicate as low as 611 and 713, respectively. Overall, the obtained coverage was in the range of other MS based studies for *P. putida*, which found
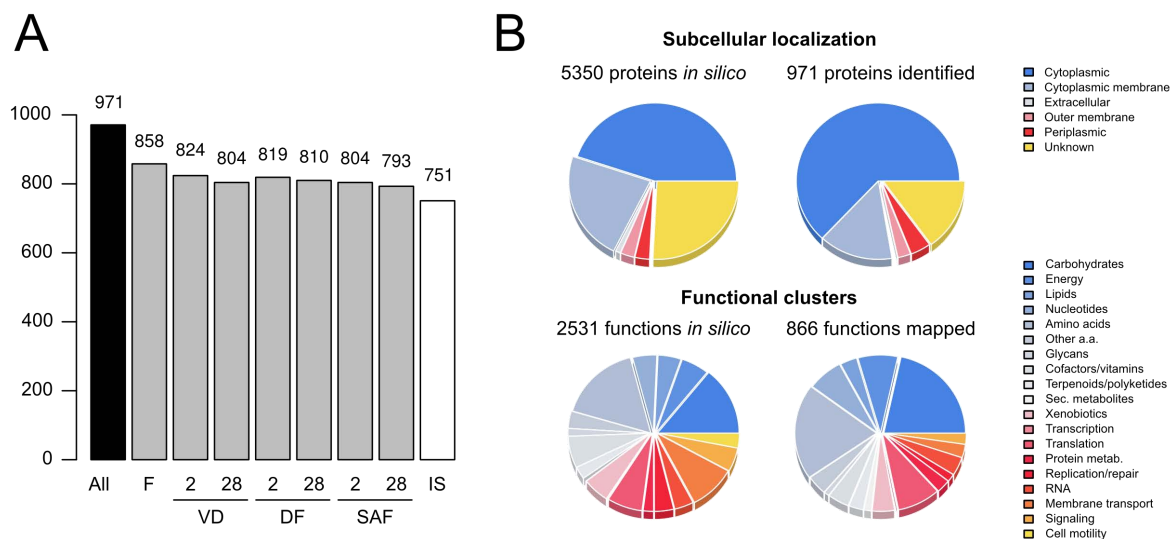
Figure 3. Proteins detected by mass spectrometry after storage of *P. putida* KT2440 cells. (**A**) Number of unique proteins present in at least one replicate per sample for fresh (F), vacuum dried (VD), deep frozen (DF) and sodium azide fixed cells (SAF). The intersection (IS) denotes the number of proteins common in all samples. (**B**) Subcellular localization of all 971 detected proteins in comparison to 5350 protein encoding sequences in the *P. putida* database (www.pseudomonas.com). A subset of the identified proteins was mapped to 866 functions by KEGG BRITE, compared to 2531 functions of all annotated proteins.

ACCEPTED ARTICLE | JIOMICS 2013

604 to 2383 different proteins [36-39]. Here, the 971 identified proteins showed a different distribution in theoretical subcellular localization (as determined *in silico* by PSORTb) compared to the 5350 proteins annotated at www.pseudomonas.com (Figure 3B). For instance, the proportion of cytoplasmic proteins was higher for the experimental data (63 % instead of 45 %), whereas the proportion of inner membrane localized proteins was smaller (15 % instead of 23 %). To gain a deeper insight into the nature of the identified proteins, 463 out of 971 unique proteins were assigned to 866 functional annotations using the KEGG BRITE database (Figure 3B), which comprised a total of 1494 unique proteins assigned to 2531 functional annotations for *P. putida* KT2440. The clustering of proteins according to the second highest hierarchy level (here called 'process') with 19 subgroups illustrates the over-representation of detected proteins in carbohydrate, energy and amino acid metabolism. An under-representation was observed for proteins involved in membrane transport, signal transduction and cell motility.

3.4. *Protein profile similarity of stored samples*

Besides the number of proteins identified per sample, the relative quantity of the detected proteins was taken into account to test the impact of the three preservation methods on cell protein composition. Therefore, we investigated if certain functional clusters of proteins were more prone to alteration than others. The hierarchical annotation by KEGG BRITE was used to arrange proteins according to the top three out of four hierarchical levels ('system', 'process' and 'pathway'). Protein clusters were drawn as a sunburst

treemap (Figure 4), where each layer of the treemap represents one hierarchical level, starting with the broadest level as innermost layer ('system') and ending with the most detailed layer at the surface ('pathway'). The width of a sector corresponds to the number of proteins within the functional group, and the color represents the standard deviation of the $\log_2$ relative protein quantity (*sd*), with the fresh sample (F) as reference. For example, a yellow or red color indicates a high *sd* and therefore a higher or lower quantity of (some) proteins in a group, compared to the proteins of the fresh sample. Thus, storage-induced variations affecting only certain functional groups can be easily spotted. The variation was small within the most groups for VD and DF samples (*sd* < 1), although the pathways for methane, glycine, pyruvate and glyoxylate metabolism as well as xenobiotics biodegradation showed increased variations ($1.2 \leq sd \leq 2.0$). The storage time of 2 or 28 d had no significant influence on the protein profile. However, a stronger overall variation was displayed by SAF treated samples already after 2 d of storage. After 28 d the variation was further increased for proteins involved in oxidative phosphorylation, protein export, ABC transporter and sulfur and lysine metabolism ($1.8 \leq sd \leq 2.2$). The variation in protein quantity within functional groups is an indicator for the impact of the preservation method. To further elucidate the cause for this variation, we analyzed the peptide coverage obtained by MS for 60 selected proteins of different chain lengths across samples (supplementary Figure S6). For these 20 largest, 20 smallest and 20 proteins of medium polypeptide chain length, the peptide coverage was very similar for all samples and no decay at the termini was observed. This finding suggests, that degradation of proteins was not the cause for the deviating
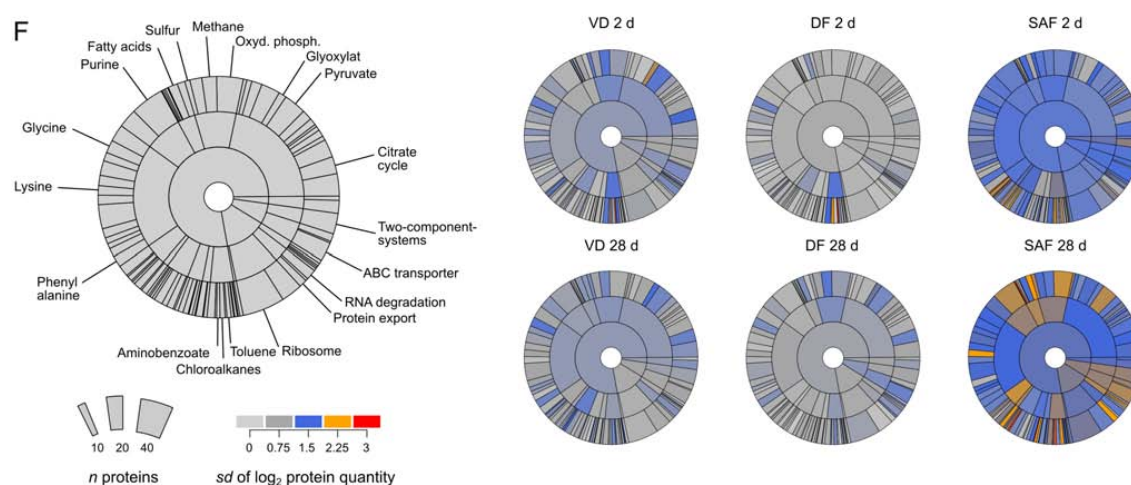
Figure 4. Functional clustering of proteins derived from cell samples prepared by VD, DF and SAF and stored for 2 d and 28 d. The sunburst treemaps represent 463 proteins mapped to functional groups using KEGG BRITE for visualization of sample variation. Treemaps consist of three hierarchically ordered layers ('system', 'process', 'pathway') with the most general ('system') being in the center. Each sector represents a group, with sector width encoding the number of proteins *n* and color encoding standard deviation of the $\log_2$ relative protein quantity (*sd*), compared to the reference (fresh cells, F). Thus, a gray color represents low variation in protein quantity and a yellow or red color a high variation.

protein profile of SAF cells in comparison to DF and VD cells. Furthermore, even a slightly increased number of peptides was detected for the SAF treated cells after 28 d compared to the 2 d stored cells. For VD, the effect was inverse with fewer peptides detected after 28 d compared to 2 d.

4. Discussion

In this study, we tested three different preservation techniques (VD, DF, SAF) for bacterial cells with respect to their influence on sample integrity. These methods were selected for speed, simplicity, and omission of organic solvents. The effect of short-term and long-term storage was compared using *P. putida* as model organism. The two techniques used here –FCM and MS– are state-of-the-art for the analysis of microbes by addressing one specific *Omics* level each, namely cytomics and proteomics. The combination of both techniques can be used to obtain a detailed picture of the cellular protein interior for different subpopulations, as was recently shown for microorganisms [2, 24]. However, this kind of experiments requires the reliable preservation of cell samples before FCM and MS are applied. Using cytometry, we found VD and DF to preserve scatter characteristics and DNA content equally well, with high similarity to fresh cells even after long term storage (28 d). Based on the results for *P. putida*, a threshold of $sd = 1.5$ can be considered appropriate for indicating high sample similarity, as measured by *FlowFP* fingerprinting (Figure 2). The similarity of SAF samples was generally lower, mainly due to differing DNA patterns of the cells.

Likewise, the protein profile of VD and DF samples for *P. putida* showed higher similarity to the fresh sample than that of SAF samples as determined by MS. Interestingly, the peptide coverage of 60 selected proteins showed no decreased numbers of detected peptides for SAF, but even a slight increase after prolonged storage (28 d). A possible cause for this may be that the three dimensional structure of proteins is disturbed or unfolded over time due to the high concentration of sodium azide salt. Depending on the nature of the used salt, cellular proteins may be more amenable for whole cell tryptic digestion under such conditions [40]. Nevertheless, the increase of peptide recovery over time is not a desirable effect, since it might influence the comparability between samples. Regarding VD and DF, the higher similarity to fresh cells in proteome profiles was also reflected on the level of peptide coverage. Most likely, the absence of an aqueous environment (VD) and the very low temperature (DF) preserves protein structure more effectively.

However, functional clustering revealed that proteins of specific pathways are more prone to alteration than others, and these pathways coincided for VD and DF (e. g. glycine and pyruvate metabolism). Similar results were found in a proteomic study with human cells, where protein degradation during cold storage affected not all proteins in a

sample and not all sample types equally [41]. Furthermore, the number of detected proteins in two of three replicates of VD and SAF was considerably lower after 28 days of storage, pointing towards the superiority of low storage temperatures (-80°C) over moderate ones (4°C). The loss of culturability of stored bacteria in relation to elevated storage temperature or time is a known phenomenon, both for freeze-dried and deep-frozen cells [15, 17, 20]. But little is known about the integrity of biomolecules in whole cells, when stored at different temperatures. What was shown at least, is the beneficial effect of deep storage temperatures (≤ -80°C) on protein stability for already isolated protein extracts [41, 42].

Most storage procedures for whole cells are optimized for biobanking, aiming at resuscitation of cells after storage. For this purpose, cryopreservation procedures like deep freezing and freeze drying may be the most important preservation methods [4]. But not every specimen is equally cryotolerant. Some microbial genera like *Helicobacter* or *Neisseria* are notoriously difficult to freeze or to recover [4] and complex microbial communities may require a completely different treatment (anaerobic sampling, metal ion treatment) for stabilization [13]. Moreover, the preservation of cells for single-cell analysis or for resuscitation are two different objectives, and the chosen technique is not necessarily suitable for both. However, cryopreservation is a preferred option, as the majority of cells is cryotolerant and the required equipment is affordable even for small laboratories [4]. The storage temperature for cryopreservation should preferably be lower than -20°C, which was reported to result in degradation of serum proteins compared to -80°C [42]. If storage capacities at -80°C are limited, alternatives like freeze drying and low temperature vacuum drying may be considered [22, 23], The vacuum drying procedure applied here preserved the cellular DNA content, light scattering properties, and protein profile with similar efficiency as deep freezing. It requires neither chemical treatment nor other technical effort than a generic vacuum concentrator, and storage of dried cells can take place at 4°C.

Whichever technique is chosen, it is necessary to test the applicability of the desired work flow to the target organism. The bacterial species covered here are not representative for all bacterial genera, but are commonly used in biotechnology. The presented preservation methods are intended for the use in sub-population proteomics, a combination of flow cytometric cell sorting and protein mass spectrometry recently applied for microbes [2, 24, 43]. As the process of cell sorting may impose further stress on recovered cells and change protein abundances independent of storage, it was omitted here. However, the results of this study may be of interest for other analytical disciplines as well. Regarding flow cytometry for instance, other markers such as fluorescent proteins are often used and their function should be conserved by DF [2]. Proteins, however, belong to the more stable biomolecules. And although more 'delicate' biomolecules such as RNA could be stronger affected by unfavorable storage conditions, the findings of this study

ACCEPTED ARTICLE | JIOMICS 2013

may very well apply to these biomolecules as well.

5. Conclusions

Three different methods for the preservation of bacterial cells –vacuum drying (VD), deep freezing (DF), sodium azide fixation (SAF)– were evaluated using flow cytometry and protein mass spectrometry. Cells of *P. putida* were stored for 2 d and 28 d and the similarity to a fresh reference sample quantified by cytometry fingerprinting and proteome profiling. Both DF and VD ensured high agreement between stored and fresh samples as analyzed by flow cytometry, whereas SAF samples showed reduced similarity. Furthermore, 971 different proteins were identified across all samples, and 463 of these proteins were functionally clustered and revealed susceptibility of certain protein groups to alteration, depending on the preservation method. Overall, most of the functional groups displayed low variation in protein quantity for VD and DF samples, but high variation in case of SAF, particularly after 28 d. Interestingly, no peptide decay — for example at the protein termini — was found for SAF but rather an increase in peptide coverage. We assume, that the protein structure is made more amenable for trypsin digestion by the action of sodium azide. Nevertheless, DF and VD are recommended for use in flow cytometry and further downstream applications like protein mass spectrometry, whereas SAF should be avoided for *P. putida*.

6. Supplementary material

Supplementary data and information is available at: http://www.jiomics.com/index.php/jio/rt/suppFiles/115/0.
Supplementary Material includes Figures S1 (Evaluation of vacuum drying (VD) conditions for *P. putida* KT2440 using flow cytometry), S2 (Evaluation of deep freezing (DF) conditions for *P. putida* KT2440 cells using flow cytometry), S3 (Internal variation of identical samples when using FlowFP fingerprinting [27]), S4 (Test of different storage methods for *E. coli* DH5α using flow cytometry), S5 (Variability of protein number and quantity across replicates as identified by mass spectrometry) and S6 (Peptide coverage of selected proteins). Supplementary material further includes S1 for MaxQuant settings, S2 for the R treemap function, and tables S3 and S4 for the list of detected proteins by MaxQuant and by Thermo Discoverer/Mascot, respectively.

Acknowledgements

References

1. B.S. Edwards, J. Zhu, J. Chen, M.B. Carter, D.M. Thal, J.J.G. Tesmer, S.W. Graves, L.A. Sklar, Cytometry A 81A (2012) 419–429. http://view.ncbi.nlm.nih.gov/pubmed/22438314
2. M. Jahn, J. Seifert, M.V. Bergen, A. Schmid, B. Bühler, S. Müller, Curr Opin Biotechnol (2012) in press. http://www.sciencedirect.com/science/article/pii/S0958166912001723
3. H.M. Shapiro, Practical flow cytometry, 4th edition, John Wiley & Sons, 2003.
4. P. De Paoli, FEMS Microbiol Rev 29 (2005) 897-910. http://view.ncbi.nlm.nih.gov/pubmed/16219511
5. D. Hopwood, Histochem J 1 (1969) 323-360. http://view.ncbi.nlm.nih.gov/pubmed/4113286
6. S. Müller, G. Nebe-von-Caron, FEMS Microbiol Rev 34 (2010) 554-587. http://view.ncbi.nlm.nih.gov/pubmed/20337722
7. M.A. Perlmutter, C.J.M. Best, J.W. Gillespie, Y. Gathright, S. González, A. Velasco, W.M. Linehan, M.R. Emmert-Buck, R.F. Chuaqui, J Mol Diagn 6 (2004) 371-377. http://view.ncbi.nlm.nih.gov/pubmed/15507677
8. S. Müller, A. Lösche, T. Bley, Acta Biotechnologica 13 (1993) 289-297. http://dx.doi.org/10.1002/abio.370130311
9. S. Müller, Cell Prolif 40 (2007) 621-639. http://view.ncbi.nlm.nih.gov/pubmed/17877606
10. F. Palmieri, M. Klingenberg, Eur J Biochem 1 (1967) 439-446. http://view.ncbi.nlm.nih.gov/pubmed/6061963
11. T. Noumi, M. Maeda, M. Futai, FEBS Lett 213 (1987) 381-384. http://view.ncbi.nlm.nih.gov/pubmed/2881810
12. J. Weber, A.E. Senior, J Biol Chem 273 (1998) 33210-33215. http://view.ncbi.nlm.nih.gov/pubmed/9837890
13. S. Günther, T. Hübschmann, M. Rudolf, M. Eschenhagen, I. Röske, H. Harms, S. Müller, J Microbiol Methods 75 (2008) 127-134. http://view.ncbi.nlm.nih.gov/pubmed/18584902
14. C. Vogt, A. Lösche, S. Kleinsteuber, S. Müller, Cytometry A 66 (2005) 91-102. http://view.ncbi.nlm.nih.gov/pubmed/16003722
15. M.Y. Lin, S.H. Kleven, Avian Dis 26 (1982) 426-430. http://view.ncbi.nlm.nih.gov/pubmed/7049151
16. T. Ahn, S.S. Kang, C. Yun, Biotechnol Lett 26 (2004) 1593-1594. http://view.ncbi.nlm.nih.gov/pubmed/15604803
17. F. Fonseca, M. Marin, G.J. Morris, Appl Environ Microbiol 72 (2006) 6474-6482. http://view.ncbi.nlm.nih.gov/pubmed/17021195
18. S. Hoefman, K. Van Hoorde, N. Boon, P. Vandamme, P. De Vos, K. Heylen, PLoS One 7 (2012) e34196. http://view.ncbi.nlm.nih.gov/pubmed/22539945
19. Y. Wang, R. Yu, C. Chou, Int J Food Microbiol 93 (2004) 209-217. http://view.ncbi.nlm.nih.gov/pubmed/15135959
20. Y. Wong, S. Sampson, W.A. Germishuizen, S. Goonesekera, G. Caponetti, J. Sadoff, B.R. Bloom, D. Edwards, Proc Natl Acad Sci U S A 104 (2007) 2591-2595. http://view.ncbi.nlm.nih.gov/pubmed/17299039
21. A. Spengler, A. Gross, H. Kaltwasser, J Clin Pathol 45 (1992) 737. http://view.ncbi.nlm.nih.gov/pubmed/1401192
22. G. Siberry, K.N. Brahmadathan, R. Pandian, M.K. Lalitha, M.C. Steinhoff, T.J. John, Bull World Health Organ 79 (2001) 43-47. http://view.ncbi.nlm.nih.gov/pubmed/11217666

23. S.A.W. Bauer, S. Schneider, J. Behr, U. Kulozik, P. Foerst, J Biotechnol 159 (2011) 351-357. http://view.ncbi.nlm.nih.gov/pubmed/21723344

24. N. Jehmlich, T. Hübschmann, M. Gesell Salazar, U. Völker, D. Benndorf, S. Müller, M. von Bergen, F. Schmidt, Appl Microbiol Biotechnol 88 (2010) 575-584. http://view.ncbi.nlm.nih.gov/pubmed/20676634

25. J. Achilles, F. Stahl, H. Harms, S. Müller, Nat Protoc 2 (2007) 2203-2211. http://view.ncbi.nlm.nih.gov/pubmed/17853877

26. A. Lemme, L. Gröbe, M. Reck, J. Tomasch, I. Wagner-Döbler, J Bacteriol 193(2011) 1863-1877. http://view.ncbi.nlm.nih.gov/pubmed/21317319

27. W.T. Rogers, H.A. Holyst, Adv Bioinformatics (2009) 193947. http://view.ncbi.nlm.nih.gov/pubmed/19956416

28. M.L. Meistrich, W. Göhde, R.A. White, J. Schumann, Nature 274 (1978) 821-823. http://view.ncbi.nlm.nih.gov/pubmed/567280

29. M. Walberg, P. Gaustad, H.B. Steen, J Microbiol Methods 35 (1999) 167-176. http://view.ncbi.nlm.nih.gov/pubmed/10192050

30. R.C. Gentleman, V.J. Carey, D.M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A.J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J.Y.H. Yang, J. Zhang, Genome Biol 5 (2004) R80. http://view.ncbi.nlm.nih.gov/pubmed/15461798

31. F. Hahne, A.H. Khodabakhshi, A. Bashashati, C. Wong, R.D. Gascoyne, A.P. Weng, V. Seyfert-Margolis, K. Bourcier, A. Asare, T. Lumley, R. Gentleman, R.R. Brinkman, Cytometry A 77 (2010) 121-131. http://view.ncbi.nlm.nih.gov/pubmed/19899135

32. E. Marco-Urrea, S. Paul, V. Khodaverdi, J. Seifert, M. von Bergen, U. Kretzschmar, L. Adrian, J Bacteriol 193 (2011) 5171-5178. http://view.ncbi.nlm.nih.gov/pubmed/21784924

33. J. Cox, M. Mann, Nat Biotechnol 26 (2008) 1367-1372. http://view.ncbi.nlm.nih.gov/pubmed/19029910

34. N.Y. Yu, J.R. Wagner, M.R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S.C. Sahinalp, M. Ester, L.J. Foster, F.S.L. Brinkman, Bioinformatics 26 (2010) 1608-1615. http://view.ncbi.nlm.nih.gov/pubmed/20472543

35. H. Ogata, S. Goto, K. Sato, W. Fujibuchi, H. Bono, M. Kanehisa, Nucleic Acids Res 27 (1999) 29-34. http://view.ncbi.nlm.nih.gov/pubmed/9847135

36. Y. Kasahara, H. Morimoto, M. Kuwano, R. Kadoya, J Microbiol Methods 91 (2012) 434-442. http://view.ncbi.nlm.nih.gov/pubmed/23022446

37. S. Yun, G.W. Park, J.Y. Kim, S.O. Kwon, C. Choi, S. Leem, K. Kwon, J.S. Yoo, J. Lee, S. Kim, S.I. Kim, J Proteomics 74 (2011) 620-628. http://view.ncbi.nlm.nih.gov/pubmed/21315195

38. D. Wijte, B.L.M. van Baar, A.J.R. Heck, A.F.M. Altelaar, J Proteome Res 10 (2011) 394-403. http://view.ncbi.nlm.nih.gov/pubmed/20979388

39. D.K. Thompson, K. Chourey, G.S. Wickham, S.B. Thieman, N.C. VerBerkmoes, B. Zhang, A.T. McCarthy, M.A. Rudisill, M. Shah, R.L. Hettich, BMC Genomics 11 (2010) 311. http://view.ncbi.nlm.nih.gov/pubmed/20482812

40. D.L. Beauchamp, M. Khajehpour, Biophys Chem 161 (2012) 29-38. http://view.ncbi.nlm.nih.gov/pubmed/22197350

41. D. Pieragostino, F. Petrucci, P. Del Boccio, D. Mantini, A. Lugaresi, M. Tiberio, M. Onofrj, D. Gambi, P. Sacchetta, C. Di Ilio, G. Federici, A. Urbani, J Proteomics 73(2010) 579-592. http://view.ncbi.nlm.nih.gov/pubmed/19666151

42. D.H. Lee, J.W. Kim, S.Y. Jeon, B.K. Park, B.G. Han, Ann Clin Lab Sci 40(2010) 61-70. http://view.ncbi.nlm.nih.gov/pubmed/20124332

43. F. Schmidt, U. Völker, Proteomics 11 (2011) 3203-3211. http://view.ncbi.nlm.nih.gov/pubmed/21710565

Subpopulation-proteomics in prokaryotic populations

Michael Jahn[1], Jana Seifert[2], Martin von Bergen[2,3], Andreas Schmid[4], Bruno Bühler[4] and Susann Müller[1]

Clonal microbial cells do not behave in an identical manner and form subpopulations during cultivation. Besides varying micro-environmental conditions, cell inherent features like cell cycle dependent localization and concentration of regulatory proteins as well as epigenetic properties are well accepted mechanisms creating cell heterogeneity. Another suspected reason is molecular noise on the transcriptional and translational level. A promising tool to unravel reasons for cell heterogeneity is the combination of cell sorting and subpopulation proteomics. This review summarizes recent developments in prokaryotic single-cell analytics and provides a workflow for selection of single cells, low cell number mass spectrometry, and proteomics evaluation. This approach is useful for understanding the dependency of individual cell decisions on inherent protein profiles.

Addresses
[1] Helmholtz Centre for Environmental Research – UFZ, Department for Environmental Microbiology, Permoserstr. 15, 04318 Leipzig, Germany
[2] Helmholtz Centre for Environmental Research – UFZ, Department for Proteomics, Permoserstr. 15, 04318 Leipzig, Germany
[3] Helmholtz Centre for Environmental Research – UFZ, Department for Metabolomics, Permoserstr. 15, 04318 Leipzig, Germany
[4] Laboratory of Chemical Biotechnology, Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Str. 66, 44227 Dortmund, Germany

Corresponding author: Müller, Susann (susann.mueller@ufz.de)

Introduction

In biotechnology, researchers are mainly working with pure prokaryotic cultures assuming that the performances of the cells might be computable and therefore controllable by applying, for example, simple Monod-kinetics. However, it was already demonstrated by many single cell based methods like flow cytometry [1[•]] and microchip technologies [2,3] that cell behavior differs even under identical micro-environmental conditions. So far, cellular diversity in clonal cell cultures was neglected or associated with gene loss, mutations, or variations in reactor environments [4,5]. Factors such as epigenetic modifications, variance in cell size owing to asymmetric cell division, or different growth and cell cycle states were also considered to be responsible for cell-to-cell differences [6] (Figure 1). Recently, the concept of cellular noise became more and more apparent interconnecting the level of transcriptional and translational noise with variations in cell decisions. Raj and van Oudenaarden defined noise as permanently changing levels of transcripts and proteins owing to random distribution of involved molecules and enzymes within the cell [7[••]]. Noise levels can be quietened by negative feedback regulation thereby cushioning cellular responses. By contrast, positive feedback acts in a multiplying way, topping basic noise levels and leading to forthcoming cell decisions [8]. Many of these phenomena are still unclear and notoriously hard to address owing to the complexity of regulatory networks and interrelations. To date, it is still impossible to define the influence of noise on the global state of prokaryotes and to understand how noise contributes to cell decision making and therefore to cell heterogeneity.

A deeper understanding of the variability in cell behavior can be obtained by analyzing as much heterogeneous cell states as possible. In this respect, gene expression analysis on the global scale is certainly a promising approach. Most proteins are relatively stable and can be analyzed with higher reliability in contrast to quickly degraded prokaryotic mRNA transcripts. Besides, proteomic approaches provide snap-shots of complex protein expression patterns that reveal deep functional information on cells. The disadvantage is that proteins may be too stable to track down noise and most detection methods are biased towards proteins of medium or high abundance, thus rare proteins may not be detected. In this review, we discuss to what extent protein analytics of single microbial cells or subpopulations already match *Omics* standards. Examples are presented that show the extent to which a bacterial population and its proteome can be resolved.

Fluorescence based single cell protein analytics

Typically, cell to cell variation in protein abundances are determined using optical technologies like various microscopic tools or flow cytometry (FC) in combination with protein marker molecules. Large numbers of cells can then be analyzed by the marker's fluorescence intensity and the result interpreted with regard to dynamic processes or sample origin [9]. To date, fluorescent markers coupled to specifically binding antibodies are the preferred option. Up to 20 different proteins within the same sample can be detected in this way [10]. An extension of

Figure 1
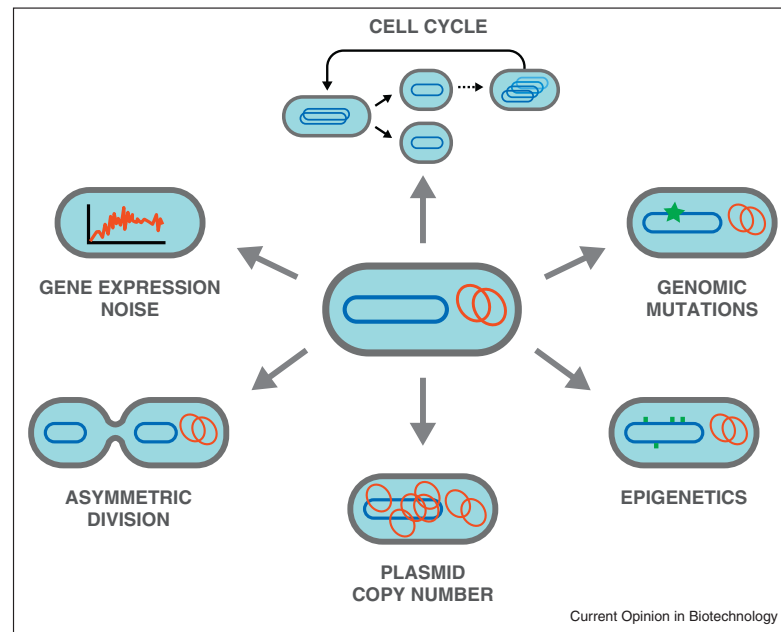
Population heterogeneity. Clonal microbial populations are heterogeneous owing to various abiotic and biotic sources acting on the single cell level.

this method is mass cytometry, which uses conjugated metal-tagged antibodies directed against surface epitopes [11••]. However, there are many pitfalls using antibodies in microbiology. First, there are only few universal antibody labeling methods available which specifically target microbial cells. Most of them act on the surface of the organisms. Binding of the probe to target proteins is often only semi-quantitative in prokaryotic cultures, especially if various cell states are investigated and multi-labeling approaches are intended. Further difficulties include quenching of adjacent dye molecules owing to small cell surfaces and volumes as well as rigid cell walls impenetrable to large probes. Similar problems occur, if amplifying fluorescent systems are used, like enzymes (horse radish peroxidase, HRP; alkaline phosphatase, AP) or TSA (tyramide signal amplification) to increase signal intensity. Furthermore, these methods require harsh and often cell wall destructive treatment of microbiological samples. Second, although a great range of antibodies (or lectins) is available, these probes are normally of meagre quality and often show unspecific binding characteristics [1•]. Third, a main tool to determine protein abundance is the use of fluorescent marker proteins to reveal, for example, oscillatory or maximum expression in regulatory circuits [12]. A wide range of fluorescent protein species can be used to quantify protein synthesis and thus track expression levels [13]. However, the function and biochemical characteristics of

the tagged proteins may be altered owing to the fusion of the fluorophore, and the equilibrium between synthesis and degradation may be affected as well [13]. In addition, engineered strains need to be used for such investigations, which require a minute genetic manipulation. Fourth, insights are constrained to the expression pattern of only one or few target molecules, when monitored by antibody binding or fluorescent protein tagging. Analysis of unique expression patterns or regulatory networks requires a more global approach.

Mass spectrometry based single cell protein analytics

Hence, we argue that *Omics* technologies will provide a deeper insight into functions of cells and reveal more complex relations to individual cell behavior and decision. However, applying proteomics to single microbial cells has not been realized to date, although different methods for analyzing various compounds of a single cell at once exist [14]. Microfluidic chips in combination with capillary electrophoresis can analyze well-chosen molecules, but do not mirror the complete cellular interior (chemical cytometry [15,16]). Microarrays, as commonly used for nucleic acids, are also used for proteins when combined with gel or mass spectrometry (MS) based tools. Related applications use antibody arrays for detection of proteins after single cell lysis on chips [17–19]. Label-free analyte measurements after cell

lysis are also described (nucleotides, chlorophyll, spiked proteins at dilutions typical for single cells [20,21]) in addition to the use of fluorescently labeled metabolites (metabolic cytometry [22]) or stable isotope markers (e.g. neuropeptides [23]). Although these techniques exhibit sensitivity sufficient for the single cell level, they do not cover the majority of proteins of a single cell and thus provide rather exclusive than global information on cellular function.

MS-based subpopulation proteomics

As the previously mentioned approaches are often premature and limited to certain pathways and regulatory circuits, researchers started to use groups of cells with similar characteristics presuming that these cells have also similar physiological characteristics. To selectively separate groups of cells from surrounding tissue or cell suspensions, laser capture micro-dissection (LCM) [24] or flow cytometric cell sorting (FS [25]) are extensively used. Using LCM, recent studies detected about 7000 proteins from pancreatic islets (each between 2000 and 4000 cells [26]) or 1700 proteins from 10 000 captured breast cancer cells [27]. Already in 2006, FS was used to separate 10^9 prokaryotic *Cupriavidus necator* cells for subsequent proteomic investigation but only 150 proteins were detected per subpopulation at the time using 2D-gel electrophoresis [28]. The large number of cells necessary to detect a considerable number of proteins points to the major problem when investigating microbial cells in comparison to eukaryotic cells: the volume factor of 1:1000 pulls down the number of detectable proteins in relation to the cell count. The combination of an improved cell sorting procedure with MS increased the resolution insofar as only 5×10^6 cells were necessary to detect about 900 proteins for *Escherichia coli* and *Pseudomonas putida* [29], limiting the time needed for sorting from 3 weeks to 3 days.

Cell selection for subpopulation proteomics

Selecting individual organisms out of populations means that subpopulations need to be defined to be selectable by LCM or FS. Microbial cells rarely show morphological characteristics that can be used to differentiate cell types within a population. Therefore, the cells need to attain fluorescence as is the case for fluorescent protein expression or be stained towards other physiologically relevant characteristics. Ideally, every cell in a population should carry or be labeled by a marker to allow not only 'yes' or 'no' decisions but also quantification. In addition to fluorescent proteins, we found DNA to be a reliable marker because every cell contains the molecule and its changing quantities mark various growth characteristics [30]. Bacteria may contain one or more chromosomes of different information and length and several copies of them, referred to as respective chromosome equivalents. DNA is a stable molecule and can be labeled

with high specificity using the blue fluorescent dye DAPI (4′,6′-diamidino-2-phenylindole) binding at A-T rich regions. It is widely discussed in microbial biotechnology that certain performances like product synthesis do only occur either when the cell cycle is finished or within the stochastic phases when no replication and cell division occurs [6]. Hence, DNA appears to be a good marker for potential decision switches in the cell cycle during recombinant protein production.

How different cells in a population can behave within identical environments is shown in the following unpublished example, based on the method described by Jehmlich *et al.* [29]. Subpopulation dynamics were analyzed for engineered *Pseudomonas putida* KT2440 during expression of the *styAB* genes encoding the styrene monooxygenase from *Pseudomonas* sp. strain VLB120 [31,32,33]. For this purpose, the cells contained a plasmid-based enhanced green fluorescent protein (EGFP) reporter construct, in which the EGFP gene was fused to the *styA* gene and gene expression was induced using dicyclopropylketone (DCPK) [34]. Besides fluorescence, cytometric analyses included DNA staining and light scattering of cells, particularly forward scattering (FSC). The growing culture (Figure 2a) showed various subpopulations regarding cell cycle stages, with cells containing C1n, C2n, C4n, and more chromosome equivalents, recombinant protein synthesis ('yes' or 'no' decisions resulting in a bimodal pattern) and light scattering (size and density). Each of the three parameters mirrored changes occurring during batch cultivation (Figure 2b). Although the DNA content of expressing and non-expressing cells was identical during exponential growth, the C1n subpopulation of expressing cells did not recover after 24 hours (Figure 2b). This hints towards impaired cell division owing to recombinant protein accumulation. The fluorescence intensity of EGFP expressing cells was already visible after 1 hour, increased to a maximum at 4 hours and slowly decreased until 24 hours after induction. EGFP is described to remain stable in cells for about 24 hours [35]. The proportion of induced cells without EGFP fluorescence increased during cultivation (1 hour: $33 \pm 10\%$, 4 hours: $35 \pm 4\%$, 8 hours: $49 \pm 11\%$, 24 hours: $55 \pm 9\%$) probably owing to the restricted cell division in expressing cells. Moreover, during the later stage of cultivation, the FSC signal of the majority of EGFP containing cells shifted towards higher intensity, most probably owing to inclusion body formation (Figure 3a and b). As a result, induced cells showed at least eight subpopulations that varied in cell abundance during the course of cultivation (Figures 2b and 3a). However, even thorough analysis of the population heterogeneity by FC does neither reveal the causes for heterogeneity nor the present state of the cells. Deeper insights into the respective subpopulations are provided if FS and MS are combined to obtain proteome data of subpopulations. One should keep in

82 **Analytical biotechnology**

Figure 2
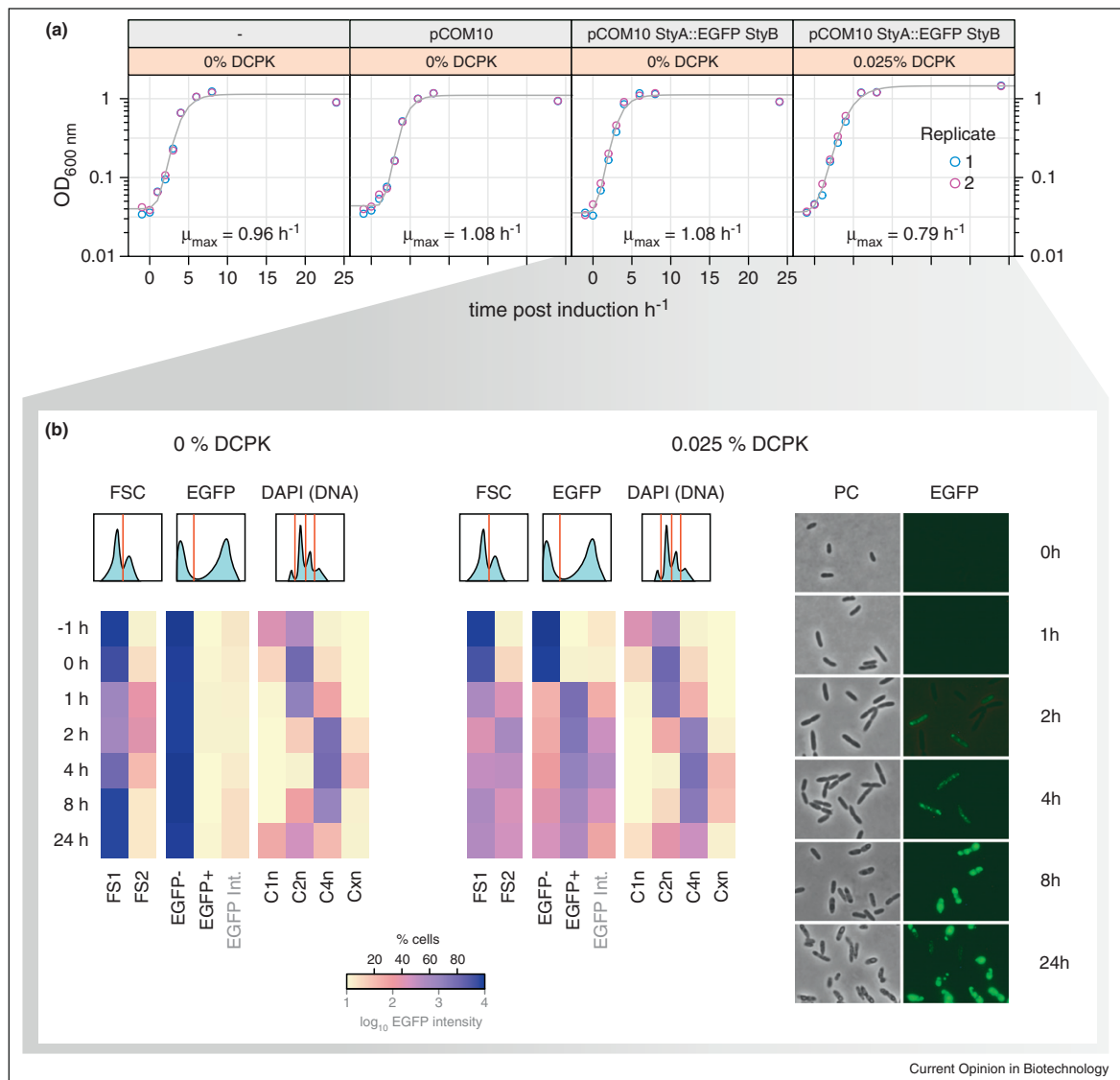
Population heterogeneity during production of a recombinant EGFP fused target protein. **(a)** Growth of wild type and recombinant *P. putida* KT2440 on citrate (5 g/L). Samples contained, from left to right, strains without plasmid (−), with target gene free plasmid (pCOM10 [34]), and with full plasmid (pCOM10 StyA∷EGFP StyB) without and with induction by 0.025% (v/v) DCPK at 0 hour. Upon induction, the synthesis of the fusion protein StyA∷EGFP resulted in decreased growth. **(b)** Heterogeneity of induced and non-induced cells in the course of protein production as determined by flow cytometry and microscopy. The proportions of distinct subpopulations differing in FSC, EGFP fluorescence, and DNA content are depicted as color-coded diagram for each gate (FS1: fluorescent cells with low FSC; FS2: fluorescent cells with high FSC; EGFP−: non-fluorescent cells; EGFP+: fluorescent cells; EGFP Int.: $\log_{10}$ fluorescence intensity; C1n: cells with one chromosome equivalent; C2n: cells with two chromosome equivalents; C4n: cells with four chromosome equivalents; Cxn: cells with more than 4 chromosome equivalents). Fluorescence and granule formation of induced cells was verified by fluorescence microscopy.

mind, that beforehand discrimination of subpopulations by specific cell characteristics, such as light scatter or fluorescence labeling, is a prerequisite for 'subpopulation-proteomics'.

MS based subpopulation proteome analysis

Subpopulation proteome analysis requires a reliable workflow which is shown in Figure 3c. In the case study using recombinant *P. putida* KT2440, only a subset of

Figure 3
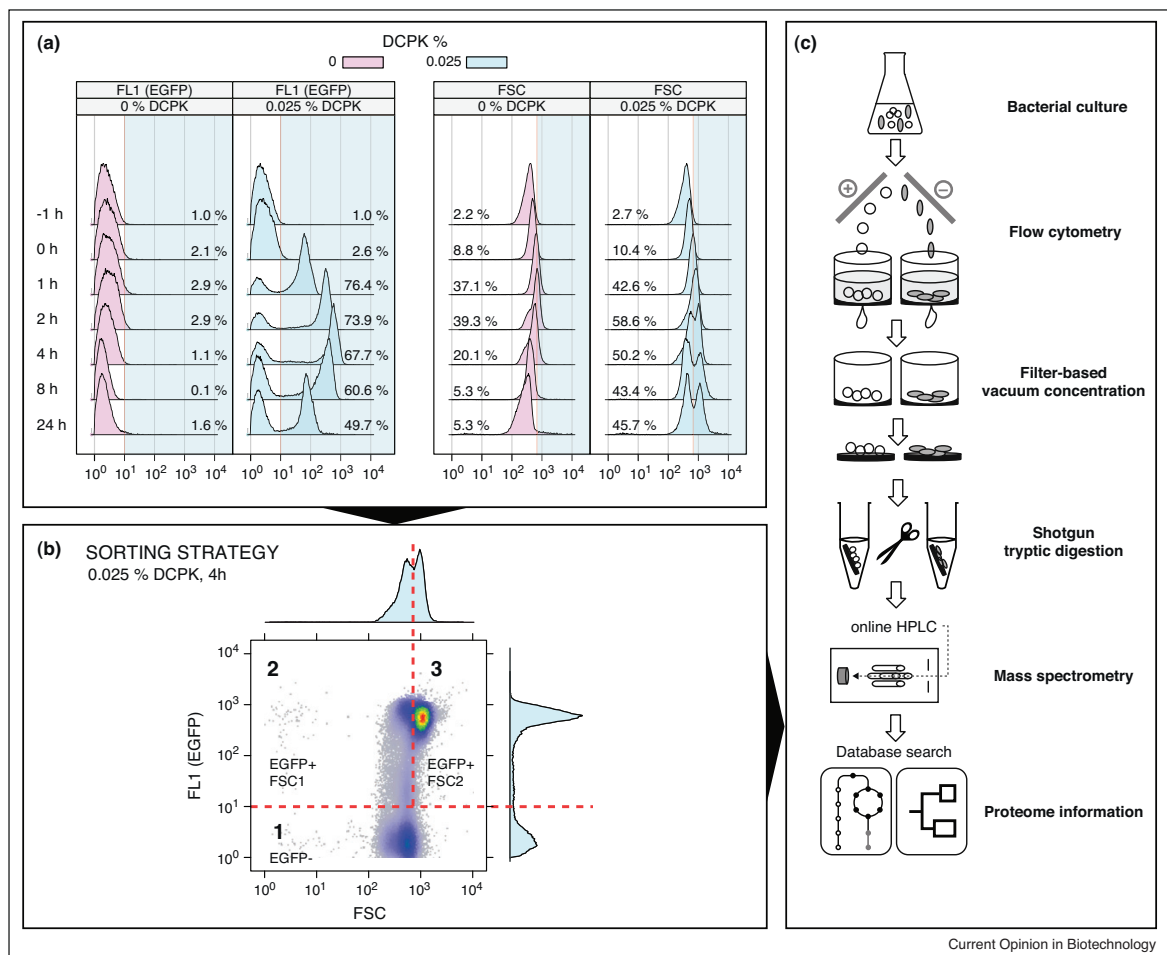
Reporter protein fluorescence and granule formation as parameters for cell sorting. **(a)** Fluorescence of the reporter protein (FL1 (EGFP)) and forward scattering (FSC) of non-induced (pink) and induced (cyan) *P. putida* KT2440 (pCOM10 StyA::EGFP StyB) as measured by flow cytometry. The induced population split in fluorescent and non-fluorescent cells, the latter was correlated with higher forward scatter owing to granule formation. **(b)** To reveal underlying changes in protein expression, cells harvested 4 hours after induction were sorted based on fluorescence and forward scatter. The gates chosen to sort the cells according to three subpopulations, being 1: cells without EGFP fluorescence, 2: cells with fluorescence and low FSC, and 3: cells with fluorescence and high FSC, are shown. **(c)** Workflow of cell sorting on filter wells, tryptic digestion, and label-free mass spectrometry for protein identification and quantification (adapted from [29]).

DCPK induced cells produced functional EGFP and these cells could be further subdivided into cells with low FSC and (more granular) cells with higher FSC. Fluorescent cells with low FSC occurred only at the beginning of cultivation (1-4 hours after induction). Therefore, the 4 hours time point was chosen for the sorting of three subpopulations, namely 1: cells without EGFP fluorescence, 2: cells with fluorescence but low FSC, and 3: cells with fluorescence and high FSC (Figure 3b). A total number of 5×10^6 cells per subpopulation with four biological replicates were sorted on a filter

well plate, trypsin digested, and analyzed by nanoLC-ESI-MS/MS (LTQ Orbitrap MS).

After identification and quantification by MaxQuant [36], a total number of 743 unique proteins (annotated and hypothetical) were detected, with 548 proteins present in at least one replicate per subproteome and 401 proteins present across all replicates. The KEGG Brite hierarchical database was used to map 373 annotated unique proteins of the dataset to 730 different functions, compared to 1235 unique entries with 2531 functions in the

hypothetical *P. putida* Brite hierarchy (Figure 4a). To find significant differences between subpopulations, two different algorithms for gene set analysis (GAGE and SAFE [37,38]) were applied. Surprisingly, they revealed that the majority of functional categories covered by KEGG was only marginally changed (Figure 4b).

The most significantly changed pathways were for instance the ones for pyrimidine and purine metabolism as well as oxidative phosphorylation. Even within such significantly changed groups, single proteins only showed small to moderate changes in expression level (Figure 4c). Beyond functional groups, the most differentially expressed genes (Figure 4c, extremes) included particularly stress related genes. In EGFP expressing cells (SP 2/SP 3), the heat shock chaperones IbpA (PP_1982) and hfq (PP_4894) as well as catalases katA (PP_0481) and ahpC (PP_2439) were found to be strongly upregulated, whereas ATPase subunit F0 (PP_5418) and elongation factor efp (PP_1858) were downregulated. Strong differences existed between subpopulations regarding the four plasmid-encoded proteins, even for the kanamycin resistance conferring KanR, although all cells were cultivated in the presence of the antibiotic (Figure 4c). Possible reasons include a reduced plasmid copy number as well as reduced expression of encoded genes.

These results suggest that the cells of subpopulation 3 accumulated the recombinant protein, ceased cell cycle activity, and activated stress response. By contrast, the non-fluorescent cells of subpopulation 1 featured increased nucleotide metabolism (DNA synthesis) and respiration despite kanamycin selection, whereas cells from subpopulation 2 represented the intermediate state of 1 and 3. This experiment may serve as an example that cytometrically analyzed intra-population variation is mirrored in the proteome profile. It is, however, remarkable that the variation between the subpopulations was restricted to a few proteins, although the high variation in optical cell characteristics implied more global differences.

Interpretation of cell-to-cell variations in clonal cultures

In the presented case study, multiple subpopulations were characterized by different cell-based and proteome-based features although environmental conditions were identical. Why do cells make these diverse decisions? And do they stick to decisions once made?

Recently, the phenomenon of noise was discussed in the context of cell decision making. Noise is described to be less perceptible in prokaryotes and less detectable on the proteome as compared to the transcriptome level [7**]. Transcriptomic changes often occur in 'bursts', which are difficult to mirror on the protein level, because of the short half-live of mRNA in bacteria contrasting with the relatively high stability of proteins [7**].

Nevertheless, proteins can be considered as the target molecules of choice to monitor cell decisions, as they are the entity defining the cell's function. To date, single-cell variability in protein expression is preferably measured by fluorescent reporter proteins, antibody-based labeling, and microfluidic chips that utilize an antibody array after single cell lysis. Although these tools are useful and convenient, the provided view on the cell state is rather limited as only predefined target proteins are considered. Subpopulation proteomics as introduced here constitute a valuable extension to the existing approaches, providing a more global view. Instead of relying on one label per experimental condition only (such as a fluorescent reporter protein), a combination of reporter proteins, functionally targeted dyes, scatter characteristics, and parameters like cell number and time can be used to discriminate subpopulations, which then can be analyzed by MS, yielding 'sub'-proteomes. Even in prokaryotes, at least several hundred proteins per subpopulation can be identified, quantified, and clustered to reconstruct pathways and regulatory networks.

By this means, only minor differences in the metabolic pathway-related protein levels were found between the three investigated subpopulations SP 1 to SP 3 of the recombinant *P. putida* strain (Figure 4c). Nevertheless, the few and pronounced effects imposed by excessive recombinant protein production, which only occurred in a certain proportion of cells, were connected to protein turnover and thus cell maintenance. Furthermore, evidence on the single-cell and subproteome level hinted towards compromised cell division of producing cells, a fact already discussed by Sevastsyanovich *et al.* [39*]. Assuming that non-producing cells keep their pace in cell cycling, they will out-compete the producers not owing to increased growth rates but simply owing to the inability of producers to divide. Interestingly, once the decision for production was made there seems to be no going back, indicating a certain inevitability in cell fate.

Cellular decision making is a highly complex process, depending on the cell's individual prerequisites and history, its micro-environment, and extrinsic and intrinsic noise. Here, as a proof-of-principle, a proteome snap-shot was provided for a certain moment in the history of the culture, but a time series of samples may allow reconstructing cellular decisions from the beginning of population splitting on. However, cell decisions in response to intrinsic noise of gene expression will be difficult to detect in this way, because effects of extrinsic and intrinsic noise can hardly be differentiated. Another way to aquire such information might be the use of advanced reporter systems [40] or techniques that resolve spatial molecule movements over time, like fluorescence correlation spectroscopy [41] or high resolution imaging in single cells [42*].

Figure 4
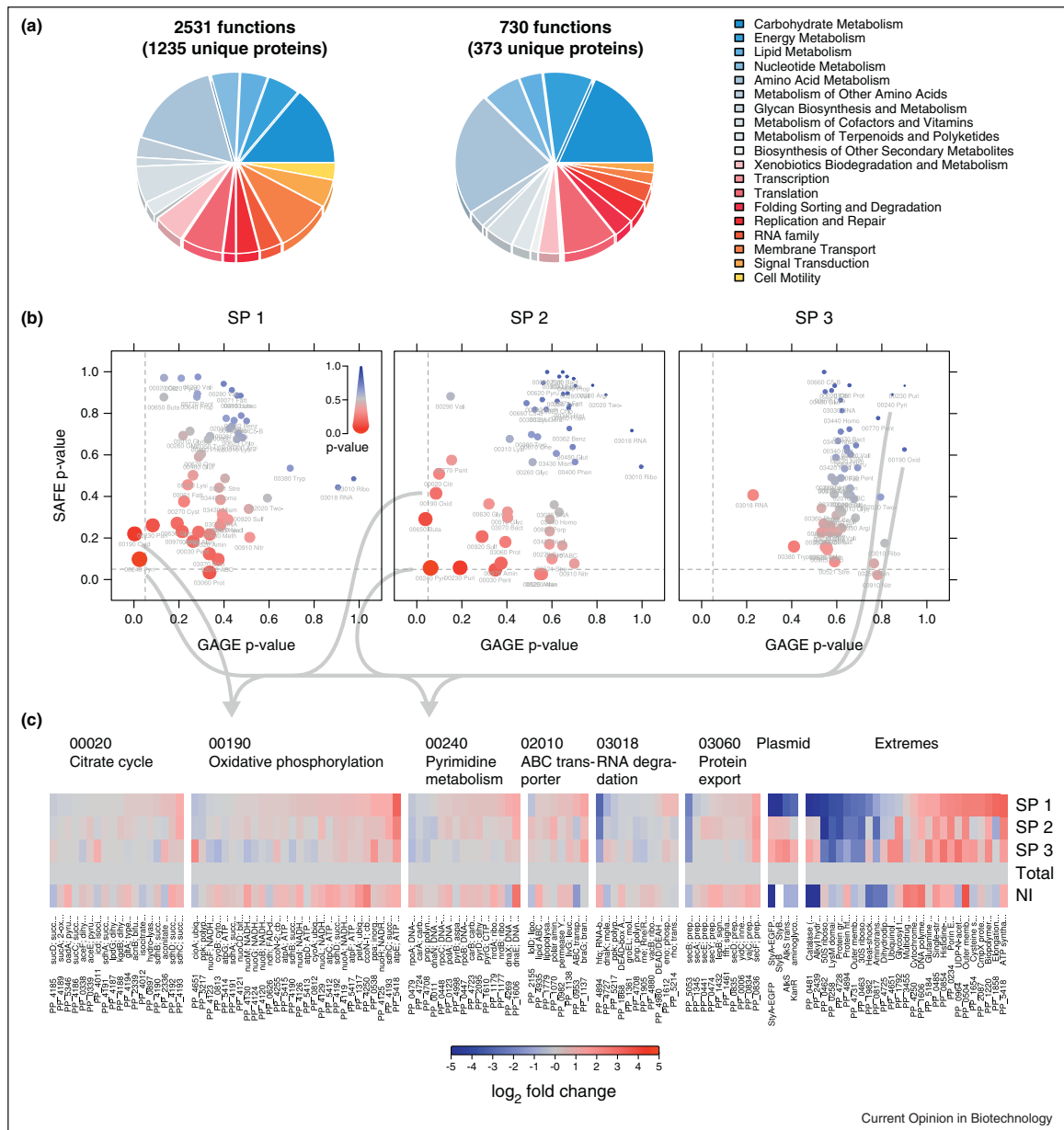
Subpopulation proteomics. Proteome data on sorted subpopulations are shown. Cells were treated as depicted in Figure 2c and subjected to mass-spectrometry with label-free quantification of proteins. **(a)** Protein identifiers mapped to functional classes using the KEGG Brite database. Indicated is the total number of assigned proteins (with number of unique proteins in brackets) for the complete proteome of *P. putida* KT2440 (left) and for the subpopulation dataset (right). **(b)** To find significant changes across KEGG pathways, gene set analysis (GAGE, SAFE) was performed for three sorted subpopulations and the *p*-values of both tests were correlated. **(c)** Relative fold change (log$_2$) of protein quantity for a selection of most significantly changed pathways. Additionally, plasmid-derived proteins and the chromosomally encoded proteins with the most extremely changed expression levels are listed without functional clustering. SP 1: cells without EGFP fluorescence; SP 2: cells with fluorescence and low FSC; SP 3: cells with fluorescence and high FSC; Total: unsorted control; NI: non-induced cells.

Bimodal distributions of markers as discussed here may emerge not only owing to one but even more decisions taking place in a cell. Strong signals like the surplus DCPK induction used in this experiment might interfere with other regulatory circuits. In turn, noisy regulation [8] may even ignore strong extrinsic signals like DCPK. As probably every gene holds its individual expression noise independent from expression level, the complex interplay of noise may generate a vibrating cell state that may lead to individual decisions.

It needs to be stated that the responsible regulatory proteins are often of small molecular mass, are low abundant and may not be detected when MS is performed on low cell numbers of sorted microbial subpopulations. Other shortcomings of this *Omics* approach are that the protein rather than the transcript level is covered, and, with cell sorting and MS usually taking several hours per sample, it requires considerable effort in time and instrumentation.

Conclusion

Specific cellular markers are required to select subpopulations for subsequent proteomic analyses. The presented method is therefore constrained to populations that show optically detectable variances in cell physiology. Emerging subpopulations can be sorted and their proteome resolved. Subproteomics allows tracking down differential gene expression in regard to cellular functions. In the presented example, we found 743 unique proteins in three subpopulations. The number of specifically regulated proteins was surprisingly low and closely connected to overexpression of the heterologous protein. Thus, using the presented workflow, microbial subpopulation proteomics showed functional variety of cells. An extension of this method by protein-SIP (protein stable-isotope probing) [43] may allow tracking of specific metabolic activity [43], and more advanced MS methods may allow the detection of very low abundant proteins.

Many cellular decisions can be followed using cell scatter characteristics, intrinsic or extrinsic fluorescent signals and subproteome analytics. However, decisions based on gene expression noise still cannot be resolved as knowledge on molecule dynamics and spatial distribution within a single prokaryotic cell is not recorded using flow cytometry. Despite some limitations, the methods presented here allow an unparalleled view into microbial population heterogeneity.

Acknowledgements

References and recommended reading
Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Müller S, Nebe-von-Caron G: **Functional single-cell analyses:**
   • **flow cytometry and cell sorting of microbial populations and communities**. *FEMS Microbiol Rev* 2010, **34**:554-587.
   The review gives an overview on the use of flow cytometry in microbiology, which includes trouble shooting steps as well as statistics.

2. Kortmann H, Blank LM, Schmid A: **Single cell analytics: an overview**. *Adv Biochem Eng Biotechnol* 2011, **124**:99-122.

3. Yasuda K: **Algebraic and geometric understanding of cells: epigenetic inheritance of phenotypes between generations**. *Adv Biochem Eng Biotechnol* 2011, **124**:55-81.

4. Schweder T, Krüger E, Xu B, Jürgen B, Blomsten G, Enfors SO, Hecker M: **Monitoring of genes that respond to process-related stress in large-scale bioprocesses**. *Biotechnol Bioeng* 1999, **65**:151-159.

5. Fritzsch FSO, Dusny C, Frick O, Schmid A: **Single-cell analysis in biotechnology, systems biology, and biocatalysis**. *Annu Rev Chem Biomol Eng* 2012, **3**:129-155.

6. Müller S, Harms H, Bley T: **Origin and analysis of microbial population heterogeneity in bioprocesses**. *Curr Opin Biotechnol* 2010, **21**:100-113.

7. Raj A, van Oudenaarden A: **Nature, nurture, or chance:**
   •• **stochastic gene expression and its consequences**. *Cell* 2008, **135**:216-226.
   The review emphasizes gene expression as a fundamentally stochastic process that causes cell-to-cell variations in mRNA and protein levels. Random fluctuations in the expression of individual genes are seen to be the source of cell variability. The phenomenon is discussed on several organismal levels and attempts are undertaken to quantify and interpret noise levels.

8. Balázsi G, van Oudenaarden A, Collins JJ: **Cellular decision making and biological noise: from microbes to mammals**. *Cell* 2011, **144**:910-925.

9. Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK: **A deep profiler's guide to cytometry**. *Trends Immunol* 2012, **33**:323-332.

10. Chattopadhyay PK, Roederer M: **Cytometry: today's technology and tomorrow's horizons**. *Methods* 2012, **57**:251-258.

11. Bendall SC, Simonds EF, Qiu P, Amir ED, Krutzik PO, Finck R,
    •• Bruggner RV, Melamed R, Trejo A, Ornatsky OI *et al.*: **Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum**. *Science* 2011, **332**:687-696.
    Single-cell atomic mass cytometry was used to simultaneously examine 34 parameters in single cells. Transition element isotopes not normally found in biological systems were used as chelated antibody tags. As the method is largely unhampered by interference from spectral overlap, it allows the detection of more parameters simultaneously than does traditional flow cytometry. The data set allowed for an algorithmically driven assembly providing tree maps of cell signaling responses.

12. Newman JRS, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS: **Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise**. *Nature* 2006, **441**:840-846.

13. Chudakov DM, Matz MV, Lukyanov S, Lukyanov KA: **Fluorescent proteins and their applications in imaging living cells and tissues**. *Physiol Rev* 2010, **90**:1103-1163.

14. Wu M, Singh AK: **Single-cell protein analysis**. *Curr Opin Biotechnol* 2012, **23**:83-88.

15. Sobhani K, Fink SL, Cookson BT, Dovichi NJ: **Repeatability of chemical cytometry: 2-DE analysis of single RAW 264.7 macrophage cells**. *Electrophoresis* 2007, **28**:2308-2313.

16. Mellors JS, Jorabchi K, Smith LM, Ramsey JM: **Integrated microfluidic device for automated single cell analysis using**

electrophoretic separation and electrospray ionization mass spectrometry**. *Anal Chem* 2010, **82**:967-973.

17. Huang B, Wu H, Bhaya D, Grossman A, Granier S, Kobilka BK, Zare RN: **Counting low-copy number proteins in a single cell**. *Science* 2007, **315**:81-84.

18. Shi Q, Qin L, Wei W, Geng F, Fan R, Shin YS, Guo D, Hood L, Mischel PS, Heath JR: **Single-cell proteomic chip for profiling intracellular signaling pathways in single tumor cells**. *Proc Natl Acad Sci USA* 2012, **109**:419-424.

19. Salehi-Reyhani A, Kaplinsky J, Burgin E, Novakova M, deMello AJ, Templer RH, Parker P, Neil MAA, Ces O, French P *et al.*: **A first step towards practical single cell proteomics: a microfluidic antibody capture chip with TIRF detection**. *Lab Chip* 2011, **11**:1256-1261.

20. Urban PL, Jefimovs K, Amantonico A, Fagerer SR, Schmid T, Mädler S, Puigmarti-Luis J, Goedecke N, Zenobi R: **High-density micro-arrays for mass spectrometry**. *Lab Chip* 2010, **10**:3206-3209.

21. Rubakhin SS, Romanova EV, Nemes P, Sweedler JV: **Profiling metabolites and peptides in single cells**. *Nat Methods* 2011, **8**:S20-S29.

22. Essaka DC, Prendergast J, Keithley RB, Palcic MM, Hindsgaul O, Schnaar RL, Dovichi NJ: **Metabolic cytometry: capillary electrophoresis with two-color fluorescence detection for the simultaneous study of two glycosphingolipid metabolic pathways in single primary neurons**. *Anal Chem* 2012, **84**:2799-2804.

23. Rubakhin SS, Sweedler JV: **Quantitative measurements of cell-cell signaling peptides with single-cell MALDI MS**. *Anal Chem* 2008, **80**:7128-7136.

24. Liu NQ, Braakman RBH, Stingl C, Luider TM, Martens JWM, Foekens JA, Umar A: **Proteomics pipeline for biomarker discovery of laser capture microdissected breast cancer tissue**. *J Mammary Gland Biol Neoplasia* 2012, **17**:155-164.

25. Di Palma S, Stange D, van de Wetering M, Clevers H, Heck AJR, Mohammed S: **Highly sensitive proteome analysis of FACS-sorted adult colon stem cells**. *J Proteome Res* 2011, **10**:3814-3819.

26. Waanders LF, Chwalek K, Monetti M, Kumar C, Lammert E, Mann M: **Quantitative proteomic analysis of single pancreatic islets**. *Proc Natl Acad Sci USA* 2009, **106**:18902-18907.

27. Thakur D, Rejtar T, Wang D, Bones J, Cha S, Clodfelder-Miller B, Richardson E, Binns S, Dahiya S, Sgroi D *et al.*: **Microproteomic analysis of 10,000 laser captured microdissected breast tumor cells using short-range sodium dodecyl sulfate-polyacrylamide gel electrophoresis and porous layer open tubular liquid chromatography tandem mass spectrometry**. *J Chromatogr A* 2011, **1218**:8168-8174.

28. Wiacek C, Müller S, Benndorf D: **A cytomic approach reveals population heterogeneity of *Cupriavidus necator* in response to harmful phenol concentrations**. *Proteomics* 2006, **6**:5983-5994.

29. Jehmlich N, Hübschmann T, Gesell Salazar M, Völker U, Benndorf D, Müller S, von Bergen M, Schmidt F: **Advanced tool for characterization of microbial cultures by combining cytomics and proteomics**. *Appl Microbiol Biotechnol* 2010, **88**:575-584.

30. Müller S: **Modes of cytometric bacterial DNA pattern: a tool for pursuing growth**. *Cell Prolif* 2007, **40**:621-639.

31. Panke S, Witholt B, Schmid A, Wubbolts MG: **Towards a biocatalyst for (S)-styrene oxide production: characterization of the styrene degradation pathway of Pseudomonas sp. strain VLB120**. *Appl Environ Microbiol* 1998, **64**:2032-2043.

32. Otto K, Hofstetter K, Röthlisberger M, Witholt B, Schmid A: **Biochemical characterization of StyAB from Pseudomonas sp. strain VLB120 as a two-component flavin-diffusible monooxygenase**. *J Bacteriol* 2004, **186**:5292-5302.

33. Blank LM, Ionidis G, Ebert BE, Bühler B, Schmid A: **Metabolic response of *Pseudomonas putida* during redox biocatalysis in the presence of a second octanol phase**. *FEBS J* 2008, **275**:5173-5190.

34. Smits TH, Seeger MA, Witholt B, van Beilen JB: **New alkane-responsive expression vectors for *Escherichia coli* and pseudomonas**. *Plasmid* 2001, **46**:16-24.

35. Andersen JB, Sternberg C, Poulsen LK, Bjorn SP, Givskov M, Molin S: **New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria**. *Appl Environ Microbiol* 1998, **64**:2240-2246.

36. Cox J, Mann M: **MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification**. *Nat Biotechnol* 2008, **26**:1367-1372.

37. Luo W, Friedman MS, Shedden K, Hankenson KD, Woolf PJ: **GAGE: generally applicable gene set enrichment for pathway analysis**. *BMC Bioinform* 2009, **10**:161.

38. Barry WT, Nobel AB, Wright FA: **Significance analysis of functional categories in gene expression studies: a structured permutation approach**. *Bioinformatics* 2005, **21**:1943-1949.

39. Sevastsyanovich YR, Alfasi SN, Cole JA: **Sense and nonsense**
• **from a systems biology approach to microbial recombinant protein production**. *Biotechnol Appl Biochem* 2010, **55**:9-28.
This review summarizes the state of the art of recombinant protein production, describes the many and common pitfalls researchers and industry face and outlines possible ways out of them.

40. Silva-Rocha R, de Lorenzo V: **A GFP-lacZ bicistronic reporter system for promoter analysis in environmental gram-negative bacteria**. *PLoS ONE* 2012, **7**:e34675.

41. Ries J, Schwille P: **Fluorescence correlation spectroscopy**. *Bioessays* 2012, **34**:361-368.

42. Perez CE, Gonzalez RLJ: **In vitro and in vivo single-molecule**
• **fluorescence imaging of ribosome-catalyzed protein synthesis**. *Curr Opin Chem Biol* 2011, **15**:853-863.
Although this review is mainly focused on translation, several interesting techniques for single-molecule imaging are presented, now rendering in-vivo observation of protein production at real-time possible.

43. Seifert J, Taubert M, Jehmlich N, Schmidt F, Völker U, Vogt C, Richnow H, von Bergen M: **Protein-based stable isotope probing (protein-SIP) in functional metaproteomics**. *Mass Spectrom Rev* 2012, **31**:683-697.

ORIGINAL ARTICLE

Open Access

Subpopulation-proteomics reveal growth rate, but not cell cycling, as a major impact on protein composition in *Pseudomonas putida* KT2440

Sarah Lieder[1], Michael Jahn[2], Jana Seifert[3,4], Martin von Bergen[3,5,6], Susann Müller[2] and Ralf Takors[1*]

Abstract

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 *Pseudomonas 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.

Keywords: Heterogeneity; Subpopulations; *Pseudomonas putida*; Proteome; Flow cytometry; Cell cycle

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 and Johnson 2004; Delvigne and Goffin 2013). 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
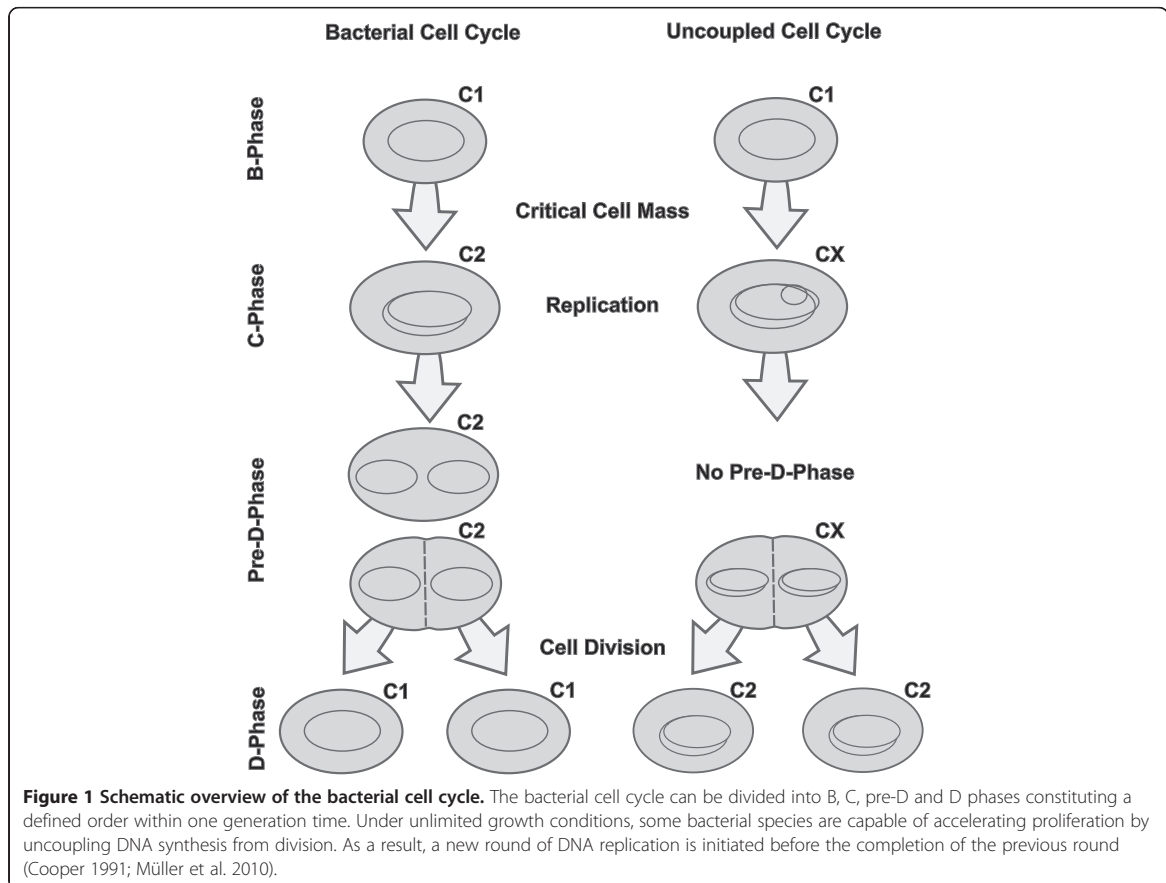
two key players in the origin of population heterogeneity, the growth rate and the cell cycle.

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 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 and Babel 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).

* Correspondence: takors@ibvt.uni-stuttgart.de
[1]Institute for Biochemical Engineering, University of Stuttgart, Allmandring 31, Stuttgart, Germany
Full list of author information is available at the end of the article

Figure 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).

It is suspected, that 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.

Pseudomonas 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 and Yokota 1973), fast growth, a fully sequenced genome (Nelson et al. 2002) and high stress tolerance (Martins

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 environmental 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 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.

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℃. Cells were cultivated in M12 minimal salt medium containing 2.2 gL^{-1} (NH$_4$)$_2$SO$_4$, 0.4 gL^{-1} MgSO$_4 \cdot$ 7 H$_2$O, 0.04 gL^{-1} CaCl$_2 \cdot$ 2 H$_2$O, 0.02 gL^{-1} NaCl, 2 gL^{-1} KH$_2$PO$_4$ and trace elements (2 mgL^{-1} ZnSO$_4 \cdot$ 7 H$_2$O, 1 mgL^{-1} MnCl$_2 \cdot$ 4 H$_2$O, 15 mgL^{-1} Na$_3$-citrate$\cdot$2 H$_2$O, 1 mgL^{-1} CuSO$_4 \cdot$ 5 H$_2$O, 0.02 mgL^{-1} NiCl$_2 \cdot$ 6 H$_2$O, 0.03 mgL^{-1} NaMoO$_4 \cdot$ 2 H$_2$O, 0.3 mgL^{-1} H$_3$BO3, 10 mgL^{-1} FeSO$_4 \cdot$ 7 H$_2$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 gL^{-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℃, a stirrer speed of 700 rpm, a pressure of 0.5 bar and an aeration of 2 Lmin^{-1} sterile filtered ambient air. The pH was set and maintained at pH 7 with 25% (v/v) NH$_4$OH. Exhaust gas composition (Blue Sense CO$_2$ and O$_2$, (DCP-CO2 DCP-02, 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 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 and Walton 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℃) HClO$_4$ solution on ice and mixed immediately (Theobald et al. 1997). The sample was shaken at 4℃ for 15 min in an overhead rotation shaker. Afterwards, the solution was neutralized on ice by fast addition of 1 mL 1 M K$_2$HPO$_4$ and 0.9 mL 5 M KOH (Buchholz et al. 2001). The neutral solution was centrifuged at 4℃ and 4,000×g for 10 min to remove cell debris, precipitated protein and potassium perchlorate. The supernatant was kept at –20℃ 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 × 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$_2$PO$_4$/K$_2$HPO$_4$, 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 and Walton (1967):

$$AEC = \frac{[ATP] + 0.5 \cdot [ADP]}{[AMP] + [ADP] + [ATP]} \tag{1}$$

Sample preparation and staining for flow cytometry

Samples for flow cytometry were washed with PBS, re-suspended 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 at 8,000×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 gL^{-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$_2$HPO$_4$), filtered through a Partec Cell-Trics mesh (Partec, Germany) with 30 μm pore size and stored on ice until analysis.

Flow cytometry and cell sorting

Flow cytometry was performed using 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. 2013; 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. 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). Cells were sorted at the most accurate mode (single cell, one drop) with a sorting speed of 4,000 s^{-1} and a sample chamber cooled to 4°C. For cell sorting a total number of 5×10^6 cells per replicate were directly sorted on a filter well plate (LoProdyne™ membrane with 0.45 μm 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 μL 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 Zip-Tip protocol (Millipore, USA), dried in a vacuum concentrator at 30°C and finally taken up in 20 μL 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).
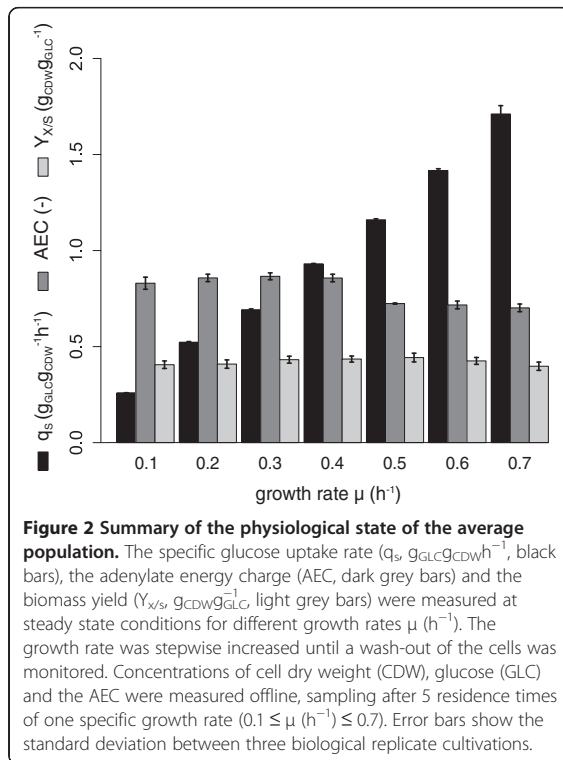
Data analysis

Mass spectra were analyzed by MaxQuant v1.2.2.5 (Cox and Mann 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 Additional file 1. The mean, standard deviation and relative quantity of replicates in relation to the reference population (RP, $\mu = 0.2$ 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) and GlobalTest (Goeman et al. 2006), setting $p < 0.05$ and a relative fold change (FC) of 1.5 ($\log_2$ FC > 0.58) as thresholds. Hierarchical groups were visualized using a color-coded circular treemap (Jahn et al. 2013).

Results

Subpopulation dynamics of *P. putida* KT2440 were analyzed in a wide range from slow growth rates starting at $\mu = 0.1$ h^{-1} to high growth rates of up to $\mu = 0.7$ h^{-1}. At growth rates higher than $\mu = 0.7$ 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, $\mu = 0.7$ 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 2). Observed stable carbon dioxide emission rates served as the criterion to qualify the achievement of steady-state cultivation conditions.

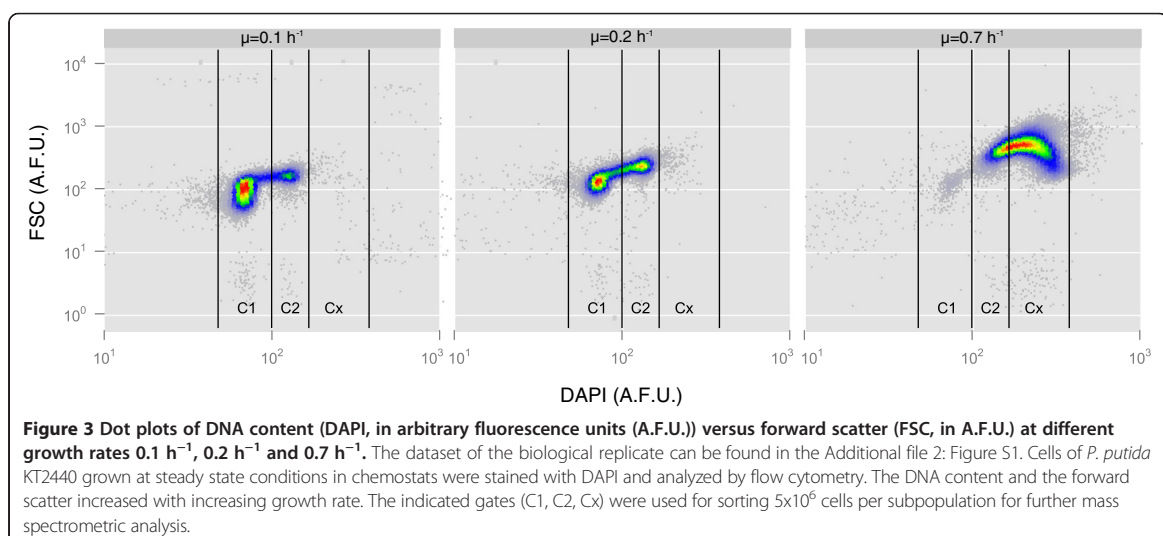The yield of biomass on glucose increased gradually by 10% from $\mu = 0.1$ h^{-1} to $\mu = 0.5$ h^{-1}. Further rise of the growth rate resulted in yield reductions, returning to the level at $\mu = 0.1$ h^{-1} (–10%). The energetic capacity of the cells can be estimated via AEC, taking the relative contribution of all three phosphorylated forms of adenine into account. The AEC was found to be stable with

Figure 2 Summary of the physiological state of the average population. The specific glucose uptake rate (q_s, $g_{GLC}g_{CDW}h^{-1}$, black bars), the adenylate energy charge (AEC, dark grey bars) and the biomass yield ($Y_{X/S}$, $g_{CDW}g_{GLC}^{-1}$, 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 ($0.1 \leq \mu$ (h^{-1}) ≤ 0.7). Error bars show the standard deviation between three biological replicate cultivations.

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 and Babel 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 and Nebe-Von-Caron 2010) (Figure 3). The dataset of the biological replicate can be found in the Additional file 2: Figure S1. 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 3). Population composition with respect to DNA content varied clearly as a function of growth rates. At $\mu = 0.1$ 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 $\mu = 0.7$ 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

increasing growth rate until $\mu = 0.4$ h^{-1}. Further increasing the growth rate resulted in a reduction of the AEC level by -18% (*p*-value < 0.01), which was almost the same at maximum growth, still staying in the range of expected physiological levels. The specific glucose uptake rate q_s was increasing linearly with increasing growth rate.

Figure 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 the Additional file 2: Figure S1. Cells of *P. putida* KT2440 grown at steady state conditions in chemostats were stained with DAPI and analyzed by flow cytometry. The DNA content and the forward scatter increased with increasing growth rate. The indicated gates (C1, C2, Cx) were used for sorting 5x10[6] cells per subpopulation for further mass spectrometric analysis.

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) (Additional file 2: Figure S2). 95.2% of the non-sorted 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 (Goeman et al. 2006; Luo et al. 2009) 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 4a). When looking at single proteins, only three were detected that comprised significantly different levels between subpopulations at growth rate $\mu = 0.1$ h^{-1} and $\mu = 0.7$ 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 and Errington 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 $\mu = 0.7$ h^{-1}, only three proteins, the translocation protein TolB (FC 1.8), the NADH dehydrogenase subunit G (PP_4124, FC 1.51) and a succinyl-diaminopimelate 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. 2006)) (Figure 4b and 4c). At $\mu = 0.1$ 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 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 ($\mu = 0.7$ h^{-1}) revealed higher presence 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 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.

Discussion

Considering the influence of different growth rates on the population, proteome analysis revealed that slow growth triggered starvation response, while fast growing cells displayed 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 ($\mu = 0.1$ 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 and Bertin 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 and Parrou 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 5). This finding is also in agreement

Figure 4 (See legend on next page.)
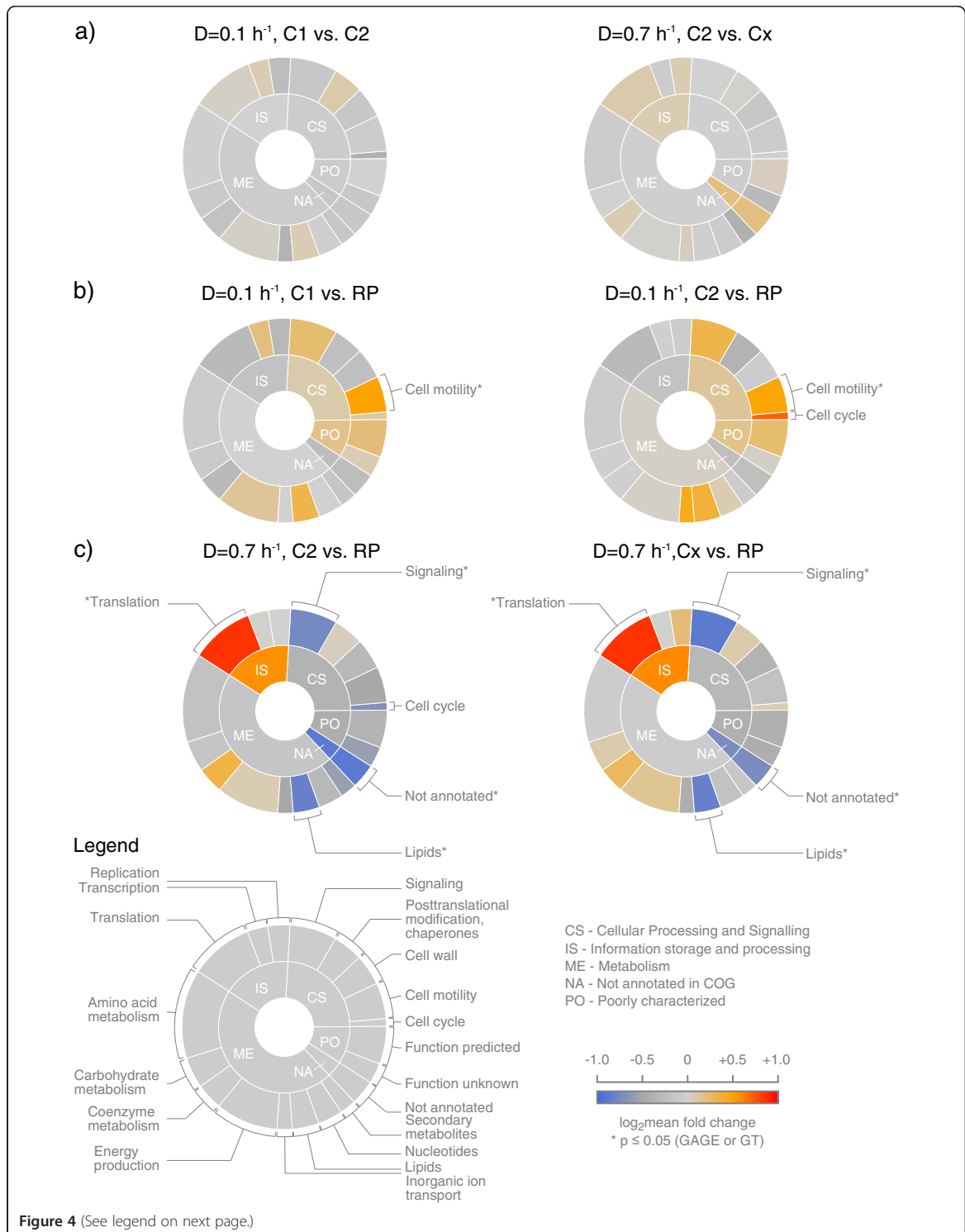
(See figure on previous page.)
Figure 4 Circular treemaps visualizing differentially expressed functional protein categories. Proteins detected by mass spectrometry were clustered according to their pathway annotation in COG covering two levels of specificity (Tatusov et al. 1997). The size of a sector 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 ($\log_2$ 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, $\mu = 0.2$ h^{-1}). Pathways with a fold change in the range $\log_2$ FC < -0.58 and $\log_2$ 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. 2006) 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}, respectively. **b.** Comparison of the subpopulations C1 and C2 at $\mu = 0.1$ h^{-1} with RP. **c.** Comparison of the subpopulations C2 and Cx at $\mu = 0.7$ h^{-1} with RP.

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 for this pathway (Aldor and Keasling 2003).

To our surprise, the almost 6.5-fold increase of the specific glucose uptake rate with increasing growth rate (Figure 2), was not mirrored by major changes

Figure 5 Heatmaps of metabolic pathways of special interest. The $\log_2$ 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 Additional file 1. One line of the heatmap represents the different subpopulations (C1, C2 and Cx) at different growth rates ($\mu = 0.1$ h^{-1}, $\mu = 0.7$ h^{-1}). Proteins of the specific pathways are shown column-wise.

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 and Dennis 1996; Maaløe and Kjeldgaard 1966; Schaechter et al. 1958). In our study, the relative cell size estimation was acquired using FSC. In accordance to various other cell cycle analyses, the FSC increased with increasing growth rates (Donachie 1968; Hewitt et al. 1999; Neumeyer et al. 2013; Skarstad et al. 1983) (Figure 3). Following the rational 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
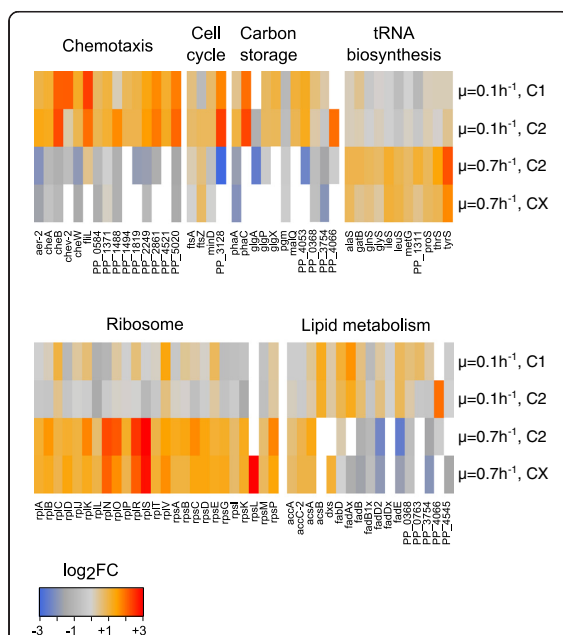
appear to be physiologically highly similar provided that the growth rate is the same.

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.

Additional files

Additional file 1: Dataset of the label-free quantification (LFQ) values that were used for further analysis of differences in protein pattern.

Additional file 2: Figure S1. Replicate dataset of dot plots of DNA content and **Figure S2**: Overview of the total protein detection and protein annotation.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
SL carried out the chemostat cultivations, analysed the datasets and drafted the manuscript. MJ carried out the flow cytometry measurements, analysed the datasets and corrected the manuscript. JS and MB performed the mass spectrometry measurements. SM and RT conceived the study and corrected the manuscript. All authors read and approved the final manuscript.

Author details
[1]Institute for Biochemical Engineering, University of Stuttgart, Allmandring 31, Stuttgart, Germany. [2]Department of Environmental Microbiology, Helmholtz Centre for Environmental Research—UFZ, Permoserstr. 15, 04318 Leipzig, Germany. [3]Department of Proteomics, Helmholtz Centre for Environmental Research—UFZ, Permoserstr. 15, 04318 Leipzig, Germany. [4]Institute of Animal Nutrition, University of Hohenheim, Emil-Wolff-Straße 8 and 10, 70599 Stuttgart, Germany. [5]Department of Metabolomics, Helmholtz Centre for Environmental Research—UFZ, Permoserstr. 15, 04318 Leipzig, Germany. [6]Department of Biotechnology, Chemistry and Environmental Engineering, University of Aalborg, Sohngaardsholmsvej 49, 9000 Aalborg, Denmark.

References
Ackermann J-U, Müller S, Lösche A, Bley T, Babel W (1995) *Methylobacterium rhodesianum* cells tend to double the DNA content under growth limitations and accumulate PHB. J Biotechnol 39:9–20

Adams DW, Errington J (2009) Bacterial cell division: assembly, maintenance and disassembly of the Z ring. Nat Rev Microbiol 7:642–653

Aldor IS, Keasling JD (2003) Process design for microbial plastic factories: metabolic engineering of polyhydroxyalkanoates. Curr Opin Biotechnol 14:475–483

Atkinson DE, Walton GM (1967) Adenosine triphosphate conservation in metabolic regulation: rat liver citrate cleavage enzyme. J Biol Chem 242:3239–3241

Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol 4:577–587

Bagdasarian M, Lurz R, Rückert B, Franklin F, Bagdasarian M, Frey J, Timmis K (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

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 J 275:5173–5190

Bley T (1990) State-structure models—A base for efficient control of fermentation processes. Biotechnol Adv 8:233–259

Brehm-Stecher BF, Johnson EA (2004) Single-cell microbiology: tools, technologies, and applications. Microbiol Mol Biol Rev 68:538–559

Bremer H, Dennis PP (1996) Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli* and *Salmonella*: cellular and molecular biology, vol 2, pp 1553–1569

Buchholz A, Takors R, Wandrey C (2001) Quantification of intracellular metabolites in *Escherichia coli* K12 using liquid chromatographic-electrospray ionization tandem mass spectrometric techniques. Anal Biochem 295:129–137

Carlquist M, Fernandes R, Helmark S, Heins A-L, Lundin L, Sorensen S, Gernaey K, Lantz A (2012) Physiological heterogeneities in microbial populations and implications for physical stress tolerance. Microb Cell Fact 11:94

Chien A-C, Hill NS, Levin PA (2012) Cell size control in bacteria. Curr Biol 22:R340–R349

Cooper S (1991) Bacterial Growth and Division: Biochemistry and Regulation of Prokaryotic and Eukaryotic Division Cycles. Academic Press, Inc., San Diego, CA

Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372

Cserjan-Puschmann M, Kramer W, Duerrschmid E, Striedner G, Bayer K (1999) Metabolic approaches for the optimisation of recombinant fermentation processes. Appl Microbiol Biotechnol 53:43–50

Delvigne F, Goffin P (2013) Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations. Biotechnol J 9:61–72

Donachie W (1968) Relationship between cell size and time of initiation of DNA replication. Nature 219:1077–1079

François J, Parrou JL (2001) Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. FEMS Microbiol Rev 25:125–145

Goeman JJ, van De Geer SA, van Houwelingen HC (2006) Testing against a high dimensional alternative. J R Stat Soc Ser B (Stat Method) 68:477–493

Harshey RM (2003) Bacterial motility on a surface: many ways to a common goal. Annu Rev Microbiol 57:249–273

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. J Biotechnol 75:251–264

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. JIOMICS 3:25–33

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. Appl Microbiol Biotechnol 88:575–584

Lee JA, Spidlen J, Boyce K, Cai J, Crosbie N, Dalphin M, Furlong J, Gasparetto M, Goldberg M, Goralczyk EM, Hyun B, Jansen K, Kollmann T, Kong M, Leif R,

McWeeney S, Moloshok TD, Moore W, Nolan G, Nolan J, Nikolich-Zugich J, Parrish D, Purcell B, Qian Y, Selvaraj B, Smith C, Tchuvatkina O, Wertheimer A, Wilkinson P, Wilson C, Wood J, Zigon R, Scheuermann RH, Brinkman RR (2008) MIFlowCyt: The minimum information about a flow cytometry experiment. Cytom Part A 73A:926–930

Lencastre Fernandes R, Nierychlo M, Lundin L, Pedersen AE, Puentes Tellez PE, Dutta A, Carlquist M, Bolic A, Schäpper D, Brunetti AC, Helmark S, Heins A-L, Jensen AD, Nopens I, Rottwitt K, Szita N, van Elsas JD, Nielsen PH, Martinussen J, Sørensen SJ, Lantz AE, Gernaey KV (2011) Experimental methods and modeling techniques for description of cell population heterogeneity. Biotechnol Adv 29:575–599

Lindmo T (1982) Kinetics of protein and DNA synthesis studied by mathematical modelling of flow cytometric protein and DNA histograms. Cell Tissue Kinet 15:197–211

Luo W, Friedman M, Shedden K, Hankenson K, Woolf P (2009) GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10:161

Maaløe O, Kjeldgaard NO (1966) Control of macromolecular synthesis: a study of DNA, RNA, and protein synthesis in bacteria. W.A. Benjamin, Inc., New York, NY

Martins dos Santos VAP, Heim S, Moore ERB, Strätz M, Timmis KN (2004) Insights into the genomic basis of niche specificity of *Pseudomonas putida* KT2440. Environ Microbiol 6:1264–1286

Meijnen J-P, de Winde JH, Ruijssenaars HJ (2008) Engineering *Pseudomonas putida* S12 for efficient utilization of D-xylose and L-arabinose. Appl Environ Microbiol 74:5031–5037

Müller S (2007) Modes of cytometric bacterial DNA pattern: A tool for pursuing growth. Cell Proliferat 40:621–639

Müller S, Babel W (2003) Analysis of bacterial DNA patterns - An approach for controlling biotechnological processes. J Microbiol Meth 55:851–858

Müller S, Nebe-von-Caron G (2010) Functional single-cell analyses: Flow cytometry and cell sorting of microbial populations and communities. FEMS Microbiol Rev 34:554–587

Müller S, Harms H, Bley T (2010) Origin and analysis of microbial population heterogeneity in bioprocesses. Curr Opin Biotechnol 21:100–113

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

Nakazawa T, Yokota T (1973) Benzoate metabolism in *Pseudomonas putida (arvilla)* mt-2: demonstration of two benzoate pathways. J Bacteriol 115:262–267

Nelson K, Weinel C, Paulsen I, Dodson R, Hilbert H, Martins dos Santos V, Fouts D, Gill S, Pop M, Holmes M (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ Microbiol 4:799–808

Neumeyer A, Hübschmann T, Müller S, Frunzke J (2013) Monitoring of population dynamics of *Corynebacterium glutamicum* by multiparameter flow cytometry. Microb Biotechnol 6:157–167

Poblete-Castro I, Becker J, Dohnt K, Santos V, Wittmann C (2012) Industrial biotechnology of *Pseudomonas putida* and related species. Appl Microbiol Biotechnol 93:2279–2290

Puchałka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins dos Santos VAP (2008) Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. PLoS Comp Biol 4(10):e1000210

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. Biotechnol J 9:511–525

Rønning ØW, Pettersen EO, Seglen PO (1979) Protein synthesis and protein degradation through the cell cycle of human NHIK 3025 cells in vitro. Exp Cell Res 123:63–72

Schaechter M, Maaløe O, Kjeldgaard N (1958) Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. J Gen Microbiol 19:592–606

Shapiro HM (2000) Microbial analysis at the single-cell level: tasks and techniques. J Microbiol Meth 42:3–16

Skarstad K, Steen HB, Boye E (1983) Cell cycle parameters of slowly growing *Escherichia coli* B/r studied by flow cytometry. J Bacteriol 154:656–662

Skarstad K, Steen HB, Boye E (1985) *Escherichia coli* DNA distributions measured by flow cytometry and compared with theoretical computer simulations. J Bacteriol 163:661–668

Soutourina OA, Bertin PN (2003) Regulation cascade of flagellar expression in Gram-negative bacteria. FEMS Microbiol Rev 27:505–523

Spidlen J, Breuer K, Rosenberg C, Kotecha N, Brinkman RR (2012) FlowRepository - A resource of annotated flow cytometry datasets associated with peer-reviewed publications. Cytom Part A 81:727–731

Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278:631–637

Teather R, Collins J, Donachie W (1974) Quantal behavior of a diffusible factor which initiates septum formation at potential division sites in *Escherichia coli*. J Bacteriol 118:407–413

Theobald U, Mailinger W, Baltes M, Rizzi M, Reuss M (1997) In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: I. Experimental observations. Biotechnol Bioeng 55:305–316

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. Biotechnol Bioeng 111:359–371

Weart R, Lee A, Chien A-C, Haeusser D, Hill N, Levin P (2007) A metabolic sensor governing cell size in bacteria. Cell 130:335–347

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. Environ Microbiol 8:165–177

analytical chemistry

# Accurate Determination of Plasmid Copy Number of Flow-Sorted Cells using Droplet Digital PCR

Michael Jahn,[†] Carsten Vorpahl,[†] Dominique Türkowsky,[†] Martin Lindmeyer,[‡] Bruno Bühler,[‡] Hauke Harms,[†] and Susann Müller*,[†]

[†]Helmholtz-Centre for Environmental Research − UFZ, Permoserstr. 15, 04318 Leipzig, Germany

[‡]Laboratory of Chemical Biotechnology, Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Str. 66, 44227 Dortmund, Germany

Ⓢ *Supporting Information*

ABSTRACT: Many biotechnological processes rely on the expression of a plasmid-based target gene. A constant and sufficient number of plasmids per cell is desired for efficient protein production. To date, only a few methods for the determination of plasmid copy number (PCN) are available, and most of them average the PCN of total populations disregarding heterogeneous distributions. Here, we utilize the highly precise quantification of DNA molecules by droplet digital PCR (ddPCR) and combine it with cell sorting using flow cytometry. A duplex PCR assay was set up requiring only 1000 sorted cells for precise determination of PCN. The robustness of this method was proven by thorough optimization of cell sorting, cell disruption, and PCR conditions. When non plasmid-harboring cells of *Pseudomonas putida* KT2440 were spiked with different dilutions of the expression plasmid pA-EGFP_B, a PCN from 1 to 64 could be accurately detected. As a proof of principle, induced cultures of *P. putida* KT2440 producing an EGFP-fused model protein by means of the plasmid pA-EGFP_B were investigated by flow cytometry and showed two distinct subpopulations, fluorescent and nonfluorescent cells. These two subpopulations were sorted for PCN determination with ddPCR. A remarkably diverging plasmid distribution was found within the population, with nonfluorescent cells showing a much lower PCN (≤1) than fluorescent cells (PCN of up to 5) under standard conditions.
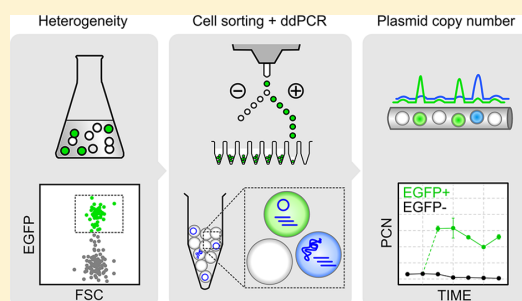
Plasmids are simple and convenient expression systems for recombinant protein production in numerous biotechnological processes.[1,2] In contrast to chromosomal expression, plasmids have the advantage of multiple copies per cell, resulting in higher gene dosage. Plasmid retention is usually ensured by means of a selection pressure, such as an antibiotic or a toxin/antitoxin system.[3] The number of plasmid copies within a cell (plasmid copy number, PCN) is a crucial factor for efficient recombinant gene expression. It is calculated as the number of plasmid copies per chromosome,[1,4] as done in this study, or alternatively as plasmid copies per cell.[5] The PCN largely depends on the plasmid's replication system and on its structural and segregational stability.[1]

Plasmid distribution is not perfectly even in a population but shows variation from cell to cell.[6] The mean and variance of PCN is determined by the segregation mechanism, which can be based on random partitioning or an active partitioning system.[7,8] Mathematical models suggest that high variance in plasmid distribution leads to more frequent occurrence of plasmid-free cells,[8] especially when random partitioning goes along with a low mean PCN.[9] The plasmid-free subpopulation can enrich and out-compete the plasmid-carrying cells.[10,11] Such processes can be monitored on the single-cell level using a

plasmid-based fluorescent reporter protein in combination with flow cytometry. For instance, in *P. putida* KT2440, plasmid-based gene expression using the *alk* regulatory system[12] was found to follow a bimodal distribution with the majority of cells having either strong fluorescence of enhanced green fluorescent protein (EGFP+) or weak enhanced green fluorescent protein (EGFP−).[13] Both cell types were sorted for proteomic analysis, which suggested either a difference in PCN or a differential expression of plasmid-located genes as the cause for heterogeneity.

Although plasmid presence can be indirectly inferred using traditional plating on selective media or fluorescent reporters,[6,13] such methods are biased by incomplete cultivability and uncontrollable influences on the formation of the reporter protein, respectively, the latter comprising variable transcription, translation, and protein folding. Direct methods to determine PCN make use of gel or capillary electrophoresis,[14,15] liquid chromatography,[16] and quantitative real-time PCR (qRT-PCR).[4,5,17] qRT-PCR proved to be specific,

accurate (CV of 10−20%[4,5]), and provided a high dynamic range of plasmid detection from 10^2 to 10^9 copies.[4] However, all these methods strongly depend on DNA standards for absolute quantification of PCN, and qRT-PCR requires the use of dilution curves for PCR efficiency calculation.[4] The PCR efficiency can dramatically influence the determination of PCN as it is the base of the Cq exponent in PCN calculation. Errors in PCR efficiency will therefore affect PCN exponentially with increasing Cq. Another source of variation is the processing of the sample to obtain analysis-grade DNA from whole cells. PCR-based methods require only small amounts of sample DNA, but extraction, purification, and linearization of DNA can distort PCN. To avoid DNA loss during preparation and simplify sample treatment, a whole cell disruption procedure using heat at 95 °C was developed for copy number analysis by qRT-PCR.[4,5] For that, the amount of starting material must be sufficient for accurate amplification without stochasticity, but on the other hand, the content of starting material should not exceed amounts where inhibition by cell debris occurs. The number of cells used per PCR reaction by Carapuca et al.[5] was at least 5×10^4, and Skulj et al.[4] used different dilutions ranging from 10^2 to 10^5 cells, the lower being the threshold for the occurrence of stochasticity.

A novel PCR-based method for highly accurate quantification of DNA copy number is digital PCR, which is based on the subdivision of a single PCR reaction mixture into many small partitions, each undergoing the PCR reaction separately.[18,19] A major advantage of digital PCR compared to qRT-PCR is the simplified experimental procedure. Neither dilution curves for PCR efficiency calculation nor DNA standards for absolute quantification are required. Digital PCR is also more precise than qRT-PCR due to higher robustness and thus allows significant detection of smaller fold changes between samples,[20−22] although qRT-PCR offers a higher dynamic range.[23] Digital PCR for DNA copy number analysis has been applied to a number of different targets,[21,22,24,25] among them plasmids from clinical[26,27] as well as environmental samples.[28] However, a thorough characterization of PCN in bacteria with focus on population heterogeneity was, to our knowledge, not performed to date.

In this study, a novel approach to determine PCN is presented using *P. putida* as a model organism. We used flow cytometry to identify subpopulations by differences in plasmid-based EGFP gene expression. A similar approach using cell sorting in combination with proteomics has already been used to analyze population heterogeneity in bacteria.[13,29,30] Here, small numbers of cells were sorted and accurately deposited in micro wells for subsequent PCN analysis by droplet digital PCR (ddPCR).[31] A workflow was established using duplex detection of genomic and plasmid DNA of only 1000 cells, revealing remarkable heterogeneity of plasmid copy number in a population of *P. putida* KT2440.

■ EXPERIMENTAL SECTION

The cultivation of bacteria, construction of plasmid, transformation of cells, and statistical methods are described in detail both in the experimental section and in Table S-1 of the Supporting Information.

Flow Cytometry and Cell Sorting. Deep-frozen samples were prepared as described in Jahn et al.[32] Briefly, samples were analyzed with a MoFlo cell sorter (Beckman-Coulter) as described in Jehmlich et al.[29] 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 EGFP fluorescence was detected in channel FL1 using blue laser excitation and a bandpass emission filter of 530/40 together with a neutral density filter of 0.3. The sheath buffer was composed as given in Koch et al.[33] with a 2-fold dilution. Prior to all measurements the instrument was adjusted with fluorescent beads[33] and EGFP expressing *P. putida* KT2440 cells as a biological standard. Data were recorded at a speed of 3000 cells/s (or lower for cell sorting) using Summit v4.3 software (Beckman-Coulter) and further analyzed using the Bioconductor framework for R.[34] Cells and beads (Fluoresbrite Bright Blue Microspheres, Ø = 0.5 μm, Polysciences) were sorted using the MoFlo cell sorter and deposited in 8-well PCR strips (G003-SF, Kisker Biotech) using the CyCLONE robotic tray. Unless otherwise stated, 1000 cells or beads (equaling to 1 μL of volume) were sorted into one tube of an 8-well strip prefilled with 7 μL of dH$_2$O at a speed of 100−200 particles/s. The most accurate sorting mode (Single Cell and One Drop mode) was used for highest purity. Cell and bead populations were gated according to the FSC, SSC, and FL1 (EGFP) signal intensity. To control for accurate sorting, *P. putida* KT2440 untransformed cells were spiked with beads prior to analysis, and a number of 1000 beads or cells were sorted as the no-template-control or no-plasmid-control, respectively. Sorted samples were immediately stored at −20 °C.

Sample Preparation for ddPCR. The heat treatment method previously reported[4,5] was adapted to extract DNA from whole sorted cells. If not stated otherwise, samples were thawed on ice, heated at 95 °C for 20 min in a Tetrad 2 thermo-cycler (Bio-Rad), and immediately cooled on ice again. The samples were briefly centrifuged at 500 × g for 3 s to remove residual liquid from the tube walls. Different incubation times for heat treatment ranging from 0 to 60 min were tested with ddPCR, and the condition with the highest obtained copy numbers was chosen for further experiments. Control samples containing only pure DNA as template were heated for 5 min. For dilution curves, samples were spiked with 2 μL of plasmid solution obtained by serial dilution of a plasmid stock (10^9 copies/μL). DNA concentration of plasmid stocks was determined using a NanoDrop spectrophotometer (Thermo Scientific).

Primer and Probe Design. Primers and probes were designed with Primer3[35] and optimized with PerlPrimer[36] regarding primer dimers, self-priming, melting temperature, aspired G/C content of 30−80%, and the presence of GC clamps. Gene sequences were retrieved from www.pseudomonas.com, and all oligonucleotides were tested for specificity using Primer-BLAST (NCBI) with the genome sequence of *P. putida* KT2440 and *E. coli* BL21 as a negative control. Designed probes were modified at the 5′ end with the fluorophore FAM for reference genes *ileS/glyA* and HEX for the plasmid-located markers *styA/oriT,* and all were modified at the 3′ end with the quencher BHQ-1. All oligonucleotides were obtained from Eurofins MWG Operon. Probes and primers were tested according to the dMIQE guidelines,[23] including optimal concentration, annealing temperature, formation of a single product (using qRT-PCR and gel electrophoresis), and discrimination of negative and positive droplets in uniplex and duplex assays (using ddPCR). A summary of used oligonucleotides is listed in Table S-2 in Supporting Information, and the dMIQE checklist is given in Table S-3 in Supporting Information.

Droplet Digital PCR. For ddPCR, a duplex reaction setup was used with simultaneous detection of a reference gene and a plasmid-located target. A single reaction volume of 20 μL contained 10 μL 2x ddPCR Supermix (Bio-Rad), 2 μL of primers (final concentration of 900 nM) and probes (final concentration of 250 nM for *ileS/glyA*, 125 nM for *styA/oriT*), and 8 μL of template solution. A master mix containing all ingredients except the template was prepared and added to the heat-treated samples in 8-well strips. The samples were thoroughly mixed, briefly centrifuged at 500 × g for 3 s, and transferred to DG8 cartridges (Bio-Rad) for droplet generation with the QX100 system (Bio-Rad) according to the manufacturer. Generated droplets were transferred to a twin.tec 96-well PCR plate (Eppendorf) and sealed for 5 s with a heat sealer (Eppendorf). The PCR reaction was performed in a Tetrad 2 thermo-cycler with the following program: 95 °C for 10 min, 40 cycles of 94 °C for 30 s and 58 °C for 60 s, 98 °C for 10 min, with a ramp rate of 2.5 °C/s. Droplets were analyzed with the QX100 droplet reader with simultaneous detection of FAM and HEX.

■ RESULTS

DNA Extraction by Heating. The quality of PCN determination strongly depends on the correct ratio of genomic DNA (gDNA) to plasmid DNA (pDNA). Every sample treatment step increases the probability of losing DNA. Therefore, we chose a sample treatment solely based on heating at 95 °C and subsequent cooling on ice. It was shown that heat treatment in a range of 5 to 20 min is effective for cell disruption.[4] To evaluate the effect of heat treatment on DNA integrity, pure plasmid DNA was heated for a duration ranging from 0 (not heated) to 20 min and analyzed by agarose gel electrophoresis (Figure 1A). For 0 and 2 min, distinct bands representing different plasmid DNA conformations[37] were found. In contrast, heat treatment for 5 to 20 min led to fragmentation of DNA indicated by smeared bands with a size ranging from 1 to 5 kb. To evaluate the effect of heat treatment on whole bacterial cells, 1000 cells of *P. putida* KT2440 carrying the plasmid pA-EGFP_B were sorted in micro wells and heated for 0 (not heated) up to 60 min. Generally, for quantification of pDNA, the gene *styA* or the origin of transfer *oriT*, and for gDNA, the genes *glyA* or *ileS* were used as markers for detection with ddPCR, respectively (Table S-2, Supporting Information). Using a duplex reaction with *styA* and *ileS*, the concentration increased for both markers up to a duration of 20 min (Figure 1B). However, prolonged heating (≥40 min) reduced the detectability of the gDNA marker by 50%, indicating a loss of template molecules due to excessive fragmentation. The PCN was calculated as the ratio of pDNA to gDNA concentration (Figure 1C), which was highly similar for samples heated for 2 to 20 min (3.8 ≤ PCN ≤ 4.2). Nonheated samples showed a strongly reduced PCN of 2, and samples heated for 40 or 60 min showed a high PCN of 8 due to loss of gDNA (Figure 1B). As a result, a heat treatment of 20 min was used for further experiments with *P. putida*, as it provided the highest concentration for the gDNA marker.

Digital PCR for PCN Analysis of Sorted Cells. To determine the PCN of selected microbial cells in heterogeneous populations, we developed a workflow combining cell sorting and digital PCR. This workflow was optimized toward simplicity featuring as few sample preparation steps as possible (Figure 2). First, small and defined cell numbers were deposited into micro wells via cell sorting and heated for 20
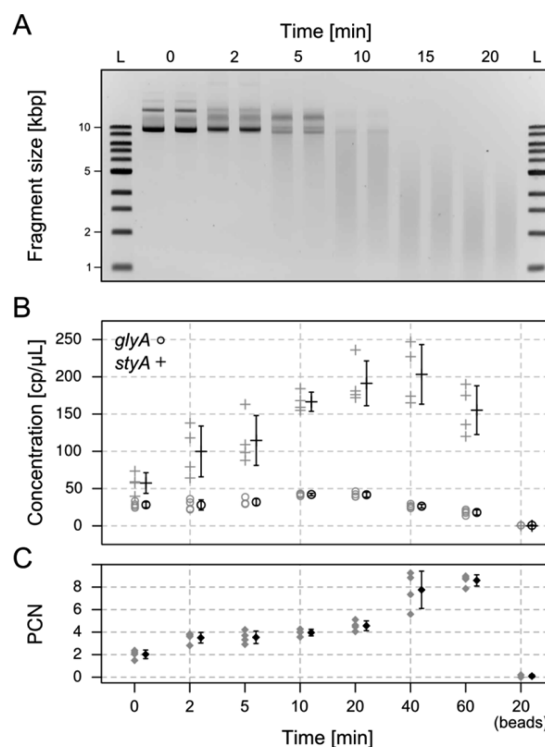

Figure 1. Heat disruption of bacteria for DNA isolation. (A) Agarose gel electrophoresis of isolated plasmid DNA (pA-EGFP_B) after heating at 95 °C for 0−20 min. kbp, kilo base pairs. L, 1 kbp ladder. (B) Digital droplet PCR for determination of plasmid copy number (PCN) using different heating times (0−60 min). The template comprised 1000 sorted cells of *P. putida* KT2440 pA-EGFP_B grown on minimal medium for 24 h. A duplex setup with one plasmid (*styA*) and one genomic DNA marker (*glyA*) was used for absolute quantification (cp/μL, copies/μL). (C) PCN was calculated as the ratio of *styA/glyA* copy number. Gray symbols, replicate values. Black symbols, mean ± standard deviation.

min at 95 °C to disrupt the cells. Then, ddPCR master mix was added, droplets were generated, the PCR reaction was performed, and the mixture of positive and negative droplets was analyzed via a microfluidic ddPCR reader. To ensure reliable DNA quantification, the PCR reaction conditions were optimized regarding primer and probe sequence, concentration, and annealing temperature.

Unlike qRT-PCR, digital PCR is more robust against a reduced PCR efficiency.[23] For example, four different primer combinations for the detection of *ileS* yielded mean fluorescence intensities of positive droplets between 4538 and 8469 arbitrary units (Figure S-1A, B, Supporting Information). However, the total number of positive droplets and thus the absolute gDNA concentration in copies/μL was highly similar for all primer pairs (37 ≤ C ≤ 44 copies/μL, Figure S-1C). Furthermore, the assay was adjusted to detect pDNA and gDNA simultaneously using a duplex setup with HEX- and FAM-labeled probes, respectively. A better discrimination of double positive droplets was achieved by reducing the concentration of HEX-labeled probes from 250 to 125 μM (Figure S-2, Supporting Information).
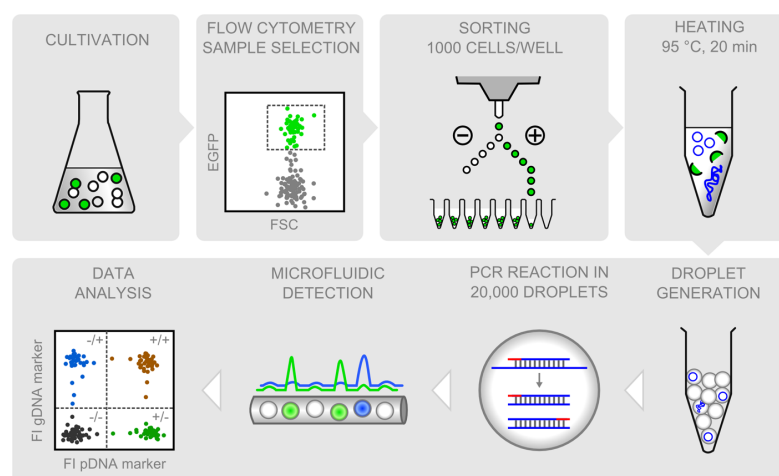
Figure 2. Workflow for cell sorting and droplet digital PCR. Cells were cultivated and analyzed by flow cytometry to detect and select subpopulations of interest (e.g., cells expressing EGFP). A number of 1000 cells per replicate was sorted in micro well tubes and heated at 95 °C for up to 20 min for cell disruption, resulting in the release of plasmid and genomic DNA (blue). PCR master mix was added to the sample (not shown), and up to 20 000 droplets were generated from a 20 μL PCR mixture. The duplex PCR reaction was performed in a standard benchtop cycler where amplification takes place in every droplet individually. Positive and negative droplets for both channels were counted by a microfluidic analyzer (gDNA marker in blue, pDNA marker in green, double positive in brown), and absolute copy number was calculated using Poisson statistics.[31]

To test the reliability of the simultaneous quantification of gDNA and pDNA, the duplex assay was compared with the uniplex assays using an artificial mixture of gDNA and pDNA as template. The respective uniplex assays yielded the same copy number for each template type as the duplex assay, excluding cross talk between both assays (Figure S-3A, Supporting Information). In a duplex PCR assay, amplification can be disturbed by different template types, such as gDNA and pDNA, interfering in the detection of each other. To exclude this, latex beads or gDNA were spiked with pDNA and used as template, resulting in no significant differences between samples (Figure S-3B).

To test the dynamic range of plasmid quantification, a $\log_2$ dilution series of pDNA was analyzed using ddPCR with 1000 sorted beads per well as a negative control for the gDNA marker (Figure 3A). The detection was linear (slope = 1.02) with a good quality of fit (R^2 = 0.992) in a range of 80 to 5000 copies/μL, spanning 7 $\log_2$ scales. In contrast to pure DNA as a template, the presence of cellular material may inhibit the PCR reaction. Therefore, the same plasmid dilution series was analyzed with a number of 1000 untransformed cells of *P. putida* KT2440 instead of beads (Figure 3B), and no significant difference between both experiments was found (slope = 0.99, R^2 = 0.996). According to this result, the influence of 1000 lysed cells on pDNA detection was negligible.

As ddPCR is a very sensitive method and theoretically able to detect a single target molecule. We tested different numbers of sorted cells ranging from 0 to 10 000 for PCN quantification. With a total number of ∼20 000 droplets per reaction, a limitation of template molecules is reasonable to avoid overloading of droplets. When no cells were sorted, ≤ 5 droplets per well were positive (0−0.5 copies/μL, Figure 4). With increasing number of sorted cells (10−1000), the precision of pDNA and gDNA quantification increased in terms of standard deviation of PCN and was optimal for 500−2000 sorted cells ($PCN_{1,000\ cells}$ = 2.2 ± 0.15). However, a cell number higher than 2000 led to reduced fluorescence for the
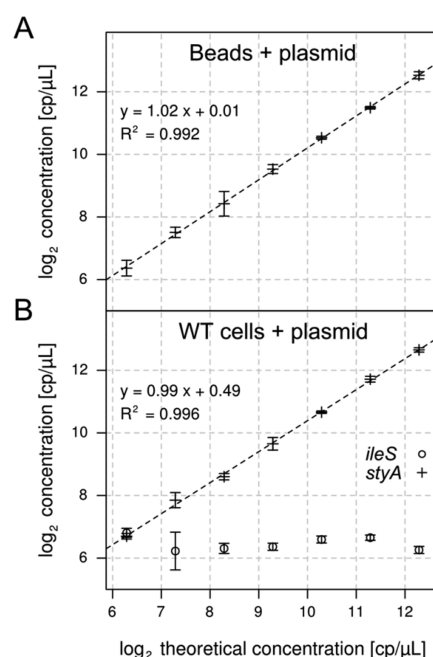

Figure 3. Plasmid DNA dilution series with sorted beads (A) and sorted cells (B) as controls. The copy number of *styA* was determined in a dilution series of plasmid pA-EGFP_B. Seven $\log_2$ dilutions of plasmid ranging from 1:1 (5000 copies/μL) to 1:64 (80 copies/μL) contained either 1000 sorted beads as no template control (A) or 1000 untransformed (WT) cells of *P. putida* KT2440 (B). Cp/μL, copies/μL.

reference gene *ileS* (Figure 4A) and therefore underestimation of gDNA copy number (Figure 4B). Thus, we used a number of 1000 sorted cells for further experiments.
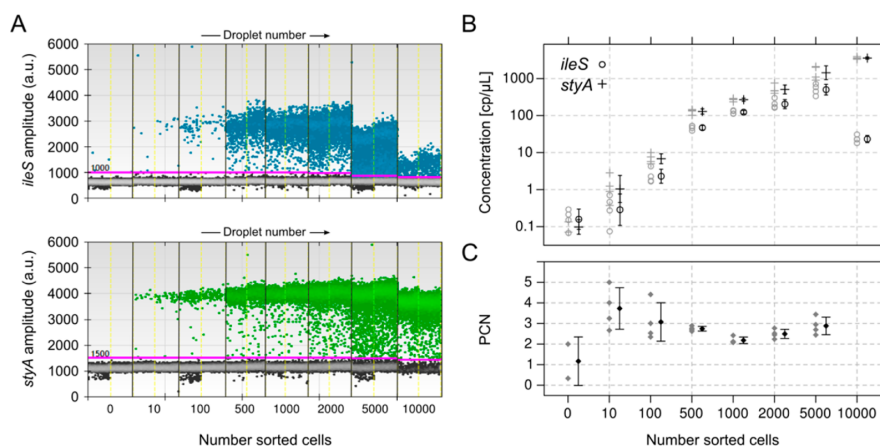
Figure 4. Different numbers of cells sorted for PCN analysis. Different cell numbers of *P. putida* KT2440 ranging from 0 to 10 000 cells were sorted, and the PCN was determined by ddPCR. (A) Dot plot of droplet fluorescence obtained via ddPCR. Two replicates of each cell number for the genomic DNA marker *ileS* and plasmid marker *styA* are shown, respectively. For high cell numbers (5000 and 10 000 cells), the fluorescence amplitude of *ileS* positive droplets declined. Arbitrary units, a.u. (B) Absolute copy numbers of *ileS* and *styA* with four replicates per condition. Cp/μL, copies/μL. (C) Plasmid copy number (PCN) of sorted cells. Gray symbols, replicate values. Black symbols, mean ± standard deviation.

Principally, it can be expected that not all selected cells are correctly sorted, lysed, and accessible for PCR. To test the ability of the assay to detect the absolute DNA copy number, cells of *P. putida* KT2440 originating from a chemostat cultivation (S. Lieder, 2014, submitted for publication) were stained with DAPI resulting in two distinct subpopulations with respect to DNA content, C1n and C2n (Figure 5A). For each of these two subpopulations, 1000 cells were sorted. If the chromosome of the C1n and C2n subpopulations contained only one and two gDNA copies each, perfect ddPCR quantification would yield 1000 and 2000 copies, respectively. The detected copy numbers in two biological replicates were 700 and 732 for C1n, and 1428 and 1396 for C2n, respectively (Figure 5B). This represents ≥70% of the theoretical copy number, showing that not all contained template molecules can be detected. Concomitantly, it was proven that 1 and 2 copies of the same gDNA marker can be discriminated in sorted cells with high precision using ddPCR and that the number of chromosomal copies (C1n, C2n) of *P. putida* KT2440 measured by flow cytometry corresponds to the number of copies measured by ddPCR.

PCN of Subpopulations during Protein Production.
The presented method was used to determine the heterogeneity of PCN in a culture of *P. putida* KT2440 producing the EGFP fusion protein StyA-EGFP. We hypothesized that EGFP− cells lost the plasmid during cultivation probably due to a naturally high resistance of *Pseudomonas* against antibiotics. To test this, cells were precultivated two times successively for 16 h in the presence of different kanamycin concentrations (0, 50, 500 μg/mL). Next, the cells were cultivated without kanamycin, as it inhibits protein synthesis at the ribosome,[38] but instead, the cells were treated with DCPK to induce StyA-EGFP protein production (Figure S-4, Supporting Information). The proportion of fluorescent cells was analyzed by flow cytometry (Figure 6A). In this analysis, 1,000 cells of the EGFP− and EGFP+ subpopulations were sorted, and the PCN was analyzed using ddPCR (Figure 6B). The EGFP+ subpopulation was first detected after 2 h of induction and increased in fluorescence intensity over a time period of 24 h.
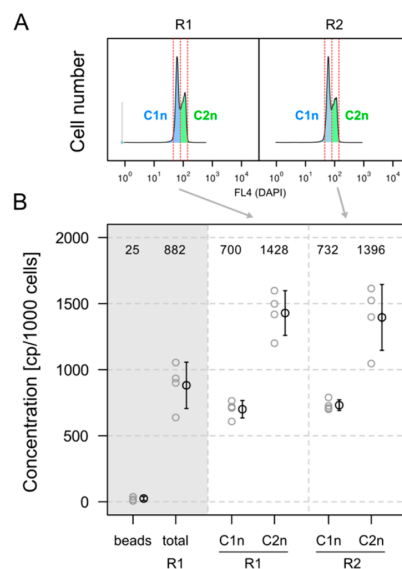

Figure 5. Genomic DNA copy number of DNA stained cells. *P. putida* KT2440 from a chemostat cultivation was stained with DAPI and analyzed by flow cytometry. (A) Histogram showing subpopulations with one and two chromosome copies (C1n, C2n). For each subpopulation, 1000 cells per well were sorted in two biological replicates (R1, R2). (B) The gDNA copy number was determined using ddPCR with the reference gene *ileS*. The obtained copies/μL were multiplied with the total reaction volume of 20 μL to yield copies/1000 sorted cells. Gray area: sorted beads as negative control, total population R1 as positive control. Gray symbols, replicate values. Black symbols, mean ± standard deviation.

The proportion of the EGFP+ cells was lowest for 0 μg/mL and highest for 500 μg/mL kanamycin in precultivation, but never exceeded 20% of all cells. The PCN of the two subpopulations was remarkably different (Figure 6B). EGFP+ cells showed a PCN of 2−3 for 0 and 50 μg/mL and a higher PCN of 3−5 for 500 μg/mL kanamycin, whereas the PCN of
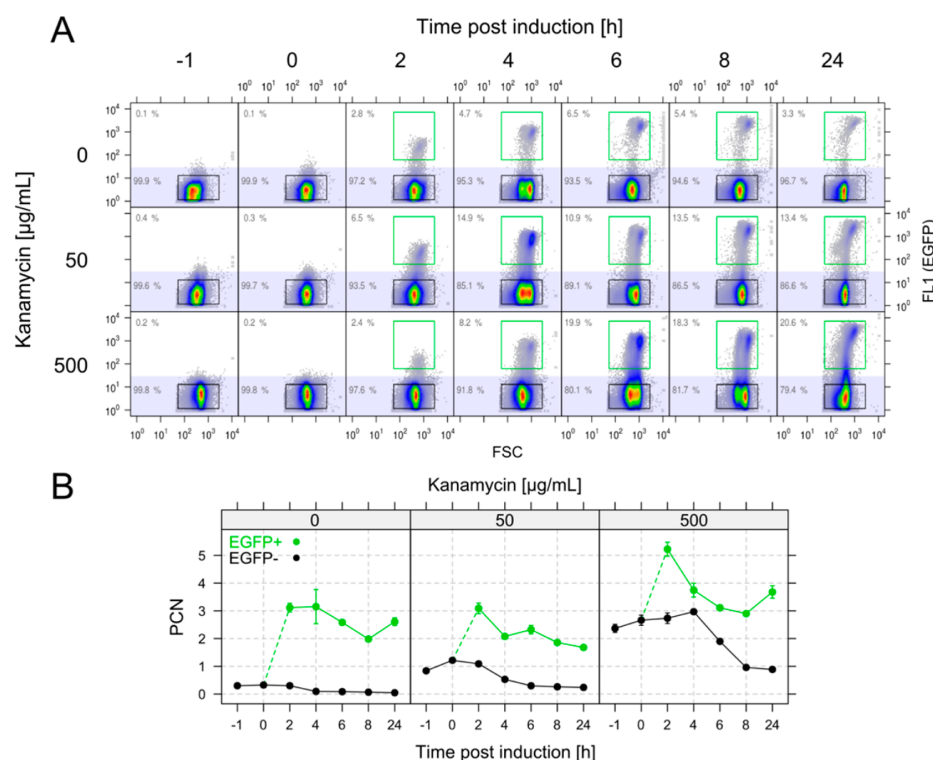
E

Figure 6. PCN after cultivation with different concentrations of antibiotic. *P. putida* KT2440 pA-EGFP_B was precultivated two times with different kanamycin concentrations (0, 50, 500 μg/mL). For protein production, a third cultivation was performed without kanamycin to prevent inhibition of translation, but DCPK was added to induce expression of StyA-EGFP. (A) StyA-EGFP formation was monitored over 24 h using flow cytometry. *X*-axis, forward scatter (FSC); *y*-axis, EGFP fluorescence. White and light blue background marks fluorescent and nonfluorescent subpopulation, respectively, with proportion of cells in gray. Rectangular gates in green (EGFP+) and black (EGFP−) surround the subpopulations that were sorted for PCN analysis. (B) PCN of sorted subpopulations (green, EGFP+; black, EGFP−).

EGFP− cells was ≤1 for 0 and 50 μg/mL. However, we found a PCN up to 3 for EGFP− cells after 4 h of cultivation when precultivated with 500 μg/mL of kanamycin. Nevertheless, for all tested conditions, the PCN of EGFP− cells decreased over time, whereas the PCN of EGFP+ cells remained rather constant. In summary, the overall plasmid copy number was low according to general classification[1] and showed strong variation between EGFP+ and EGFP− subpopulations. Under low selection pressure of 0−50 μg/mL kanamycin, the plasmid was readily lost during successive cultivation, whereas high selection pressure led to better retention of the plasmid. Notably, not all plasmid-carrying cells produced the EGFP fusion protein. It should be noted, however, that in this study, the PCN was calculated as per chromosome, and as one cell can have multiple copies of chromosomal DNA, the PCN per cell can be considerably higher.

DISCUSSION

To determine the distribution of PCN in a population, we developed a novel method based on single-cell sorting of bacteria and digital PCR. Conventional qRT-PCR and digital PCR have already been used to estimate PCN[4,5,26,27] but only for total populations of bacteria. As the mean PCN of a population conceals cells without a plasmid, we aimed to determine the PCN in different subpopulations of interest, such as protein-producing and nonproducing cells. To this end, we

used flow cytometry for cell sorting and droplet digital PCR for simultaneous detection of plasmid and genomic DNA in sorted cells. The method was applied to a *P. putida* model system with highly variable protein production, where the level of EGFP fluorescence indicates the presence or absence of the target protein.[13]

The first step of the workflow after cell sorting is DNA extraction, which was implemented by using a heat treatment, thereby omitting further washing and precipitation of DNA.[4] Heat disruption is simple, fast, and especially suitable for a small number of cells, as constituents of a greater number of lysed cells may inhibit PCR.[39] Indeed, inhibition was observed when more than 5000 cells were used in a 20 μL PCR reaction volume, which may either be due to remaining cell constituents or a high salt concentration[40] (up to 35 μM NaCl) originating from the sheath fluid in cell sorting. The optimal number of sorted cells to avoid stochasticity and inhibition ranged between 500 and 2000 cells, which is less or equal compared to other studies using whole cells for quantitative PCR.[4,5]

Heat disruption of cells is a critical step in PCN determination, because DNA is degraded by heat if not protected by (PCR) buffer.[41] We tested if heat treatment reduces DNA integrity, and we found fragmentation of plasmid DNA with increasing incubation time, as analyzed by gel electrophoresis. Nevertheless, the heat treatment was beneficial for DNA copy number yield in a range of 2 to 20 min for

P. putida KT2440, a similar range as reported for *E. coli* (10−20 min).[4,5] However, we recommend optimizing the duration of cell disruption for every species or cell type individually, as unequal degradation of plasmid and reference gene will alter the PCN.

The next step after isolation is usually the quantification of plasmid DNA, often performed with HPLC or electrophoresis. Unfortunately, both of these methods require relatively large amounts of isolated DNA and comparison with standard curves.[1] In contrast, qRT-PCR is able to detect much smaller amounts of the specimen, but it can result in false estimation of PCN when performed without reference genes (absolute quantification). One reason is that amplification of template DNA is never 100% efficient.[41] Another reason is that absolute quantification of plasmid copies per cell requires the cell number as a reference,[5] which is often indirectly and inaccurately inferred from optical density.[42,43] In contrast, the method presented here is based on sorting whole cells and performing duplex digital PCR and therefore does not suffer from these disadvantages. First of all, cell sorting allows the exact deposition of a number of cells from subpopulations of interest. For instance, we sorted cells according to their defined DNA content (C1n, C2n) and were able to distinguish one and two chromosome copies with excellent precision. Moreover, 70% of all theoretically contained DNA copies of lysed cells were detected. Nevertheless, the fact that not all theoretically contained DNA copies were detected might be due to loss of cells during sorting, although this appears improbable because a high sorting efficiency with ≥95% recovery was usually achieved with fluorescent beads. More likely, loss of DNA due to heat fragmentation or incomplete lysis of cells led to false negative droplets in ddPCR. For PCN determination, these errors can be reduced by relative quantification with a chromosomal marker as the reference. In combination with a duplex reaction setup, this also corrects well-to-well variations of cell number or reaction volume. Finally, the use of reference genes also corrects for melting double-stranded DNA to single DNA strands, which are independently distributed in a ddPCR reaction volume and hence increase copy number.[41]

Compared to qRT-PCR, digital PCR has a lower dynamic range of quantification.[23] For the purpose of PCN quantification, we tested a dynamic range of 1 to 64 plasmid copies per chromosome and verified that quantification was highly linear and accurate. The determination of higher PCNs may be possible as well. But the bottleneck of digital PCR is the number of partitions (droplets) per reaction[23] and, to avoid inhibition or overloading of droplets, a reduction of cell number to less than 1000 could be a solution, as evaluated in this study. If, however, ideal conditions without inhibition are assumed, the total number of 20 000 droplets would lead to an upper detection limit of around 10^5 template molecules, corresponding, for example, to 10^4 cells with a PCN of 10, whereas higher cell numbers would result in saturation (positive droplets only). In comparison, qRT-PCR was reported to be linear in a higher dynamic range of $10^2–10^5$ bacteria with a PCN of up to 250.[4] Apart from this limitation, digital PCR is more robust regarding systematic errors when compared with qRT-PCR, as PCR efficiency does not play a role for quantification.

As an example, the PCN was determined in a highly heterogeneous population of *P. putida* KT2440 producing the plasmid-borne recombinant protein StyA-EGFP. With EGFP fluorescence as a signal for protein production, a strongly reduced PCN was found in EGFP− cells compared to EGFP+ cells, when precultivated under standard selection conditions (50 μg/mL kanamycin). Thus, the missing protein production can be directly ascribed to the lack of the respective genetic information. But interestingly, when the concentration of antibiotic was raised 10-fold, the PCN of both subpopulations, EGFP+ and EGFP− cells, increased—some cells carried the plasmid, but were nevertheless unable to produce the desired protein. This implies that impaired protein production was not solely caused by plasmid loss but by other factors as well. The formation of functional EGFP may be slowed down because it is fused to StyA and requires the time-consuming maturation of the chromophore.[44,45] Conversely, the full induction leads to deleterious effects on translation or viability of cells.[11] Lastly, the antibiotic kanamycin inhibits protein synthesis directly at the ribosome.[38] Although it was not added during induction, it is possible that kanamycin still remains bound to the ribosome due to its high binding affinity ($1.4 \times 10^{-6} \leq K_d \leq 1.8 \times 10^{-5}$),[46−49] or leads to aggregation of misfolded proteins during precultivation,[50] and thereby stresses cells even after the removal of the antibiotic.

CONCLUSION

We developed a method for the quantification of plasmid copy number in bacteria using accurate sorting of intact cells by flow cytometry, heat disruption of cells, and quantification of plasmid and genomic DNA markers by droplet digital PCR. This workflow allowed the precise and easy analysis of PCN using only 1000 sorted cells and provided new insights in unequal plasmid distribution as a driver for population heterogeneity. Using a model gene expression system in *P. putida* KT2440, the heterogeneity in protein production could be directly related to plasmid copy number variation, although other reasons preventing the formation of EGFP may play a role as well. Apart from plasmid copy number, the combination of cell sorting and digital PCR can be used to quantify any DNA molecule or sequence in question and may even be adapted for RNA quantification.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Contains experimental details, tables listing the oligonucleotide primers for construction of plasmid pA-EGFP_B, primers and probes for ddPCR, and the dMIQE checklist. Contains figures describing the influence of primer combination and probe dilution on ddPCR, assay and template controls for ddPCR, and the cultivation procedure for *P. putida* KT2440. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author
*E-mail: susann.mueller@ufz.de. Tel.: +49 341 235 1318.

Author Contributions
The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes
The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Analytical Chemistry

■ REFERENCES

(1) Friehs, K. *Adv. Biochem. Eng./Biotechnol.* **2004**, *86*, 47−82.
(2) Vary, P. S.; Biedendieck, R.; Fuerch, T.; Meinhardt, F.; Rohde, M.; Deckwer, W.; Jahn, D. *Appl. Microbiol. Biotechnol.* **2007**, *76*, 957−967.
(3) Kroll, J.; Klinter, S.; Schneider, C.; Voss, I.; Steinbüchel, A. *Microb. Biotechnol.* **2010**, *3*, 634−657.
(4) Skulj, M.; Okrslar, V.; Jalen, S.; Jevsevar, S.; Slanc, P.; Strukelj, B.; Menart, V. *Microb. Cell Fact.* **2008**, *7*, 6.
(5) Carapuça, E.; Azzoni, A. R.; Prazeres, D. M. F.; Monteiro, G. A.; Mergulhão, F. J. M. *Mol. Biotechnol.* **2007**, *37*, 120−126.
(6) Wong Ng, J.; Chatenay, D.; Robert, J.; Poirier, M. G. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2010**, *81*, 011909.
(7) Reyes-Lamothe, R.; Nicolas, E.; Sherratt, D. J. *Annu. Rev. Genet.* **2012**, *46*, 121−143.
(8) Summers, D. K. *Trends Biotechnol.* **1991**, *9*, 273−278.
(9) Kim, B. G.; Shuler, M. L. *Biotechnol. Bioeng.* **1991**, *37*, 1076−1086.
(10) Corchero, J. L.; Villaverde, A. *Biotechnol. Bioeng.* **1998**, *58*, 625−632.
(11) Sevastsyanovich, Y.; Alfasi, S.; Overton, T.; Hall, R.; Jones, J.; Hewitt, C.; Cole, J. *FEMS Microbiol. Lett.* **2009**, *299*, 86−94.
(12) Smits, T. H.; Seeger, M. A.; Witholt, B.; van Beilen, J. B. *Plasmid* **2001**, *46*, 16−24.
(13) Jahn, M.; Seifert, J.; von Bergen, M.; Schmid, A.; Bühler, B.; Müller, S. *Curr. Opin. Biotechnol.* **2013**, *24*, 79−87.
(14) Breuer, S.; Marzban, G.; Cserjan-Puschman, M.; Dürrschmid, E.; Bayer, K. *Electrophoresis* **1998**, *19*, 2474−2478.
(15) Projan, S. J.; Carleton, S.; Novick, R. P. *Plasmid* **1983**, *9*, 182−190.
(16) Coppella, S. J.; Acheson, C. M.; Dhurjati, P. *J. Chromatogr.* **1987**, *402*, 189−199.
(17) Lee, C. L.; Ow, D. S. W.; Oh, S. K. W. *J. Microbiol. Methods* **2006**, *65*, 258−267.
(18) Sykes, P. J.; Neoh, S. H.; Brisco, M. J.; Hughes, E.; Condon, J.; Morley, A. A. *BioTechniques* **1992**, *13*, 444−449.
(19) Vogelstein, B.; Kinzler, K. W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9236−9241.
(20) Hindson, C. M.; Chevillet, J. R.; Briggs, H. A.; Gallichotte, E. N.; Ruf, I. K.; Hindson, B. J.; Vessella, R. L.; Tewari, M. *Nat. Methods* **2013**, *10*, 1003−1005.
(21) Strain, M. C.; Lada, S. M.; Luong, T.; Rought, S. E.; Gianella, S.; Terry, V. H.; Spina, C. A.; Woelk, C. H.; Richman, D. D. *PLoS One* **2013**, *8*, e55943.
(22) Whale, A. S.; Huggett, J. F.; Cowen, S.; Speirs, V.; Shaw, J.; Ellison, S.; Foy, C. A.; Scott, D. J. *Nucleic Acids Res.* **2012**, *40*, e82.
(23) Huggett, J. F.; Foy, C. A.; Benes, V.; Emslie, K.; Garson, J. A.; Haynes, R.; Hellemans, J.; Kubista, M.; Mueller, R. D.; Nolan, T.; Pfaffl, M. W.; Shipley, G. L.; Vandesompele, J.; Wittwer, C. T.; Bustin, S. A. *Clin. Chem.* **2013**, *59*, 892−902.
(24) Whale, A. S.; Cowen, S.; Foy, C. A.; Huggett, J. F. *PLoS One* **2013**, *8*, e58177.
(25) Miotke, L.; Lau, B. T.; Rumma, R. T.; Ji, H. P. *Anal. Chem.* **2014**, *86*, 2618−2624.
(26) Last, A. R.; Roberts, C. H.; Cassama, E.; Nabicassa, M.; Molina-Gonzalez, S.; Burr, S. E.; Mabey, D. C. W.; Bailey, R. L.; Holland, M. J. *J. Clin. Microbiol.* **2014**, *52*, 324−327.
(27) Straub, T.; Baird, C.; Bartholomew, R. A.; Colburn, H.; Seiner, D.; Victry, K.; Zhang, L.; Bruckner-Lea, C. J. *J. Microbiol. Methods* **2013**, *92*, 9−10.
(28) Dong, L.; Meng, Y.; Wang, J.; Liu, Y. *Anal. Bioanal. Chem.* **2014**, *406*, 1701−1712.
(29) Jehmlich, N.; Hübschmann, T.; Gesell Salazar, M.; Völker, U.; Benndorf, D.; Müller, S.; von Bergen, M.; Schmidt, F. *Appl. Microbiol. Biotechnol.* **2010**, *88*, 575−584.
(30) Wiacek, C.; Müller, S.; Benndorf, D. *Proteomics* **2006**, *6*, 5983−5994.
(31) Hindson, B. J.; Ness, K. D.; Masquelier, D. A.; Belgrader, P.; Heredia, N. J.; Makarewicz, A. J.; Bright, I. J.; Lucero, M. Y.; Hiddessen, A. L.; Legler, T. C.; Kitano, T. K.; Hodel, M. R.; Petersen, J. F.; Wyatt, P. W.; Steenblock, E. R.; Shah, P. H.; Bousse, L. J.; Troup, C. B.; Mellen, J. C.; Wittmann, D. K.; Erndt, N. G.; Cauley, T. H.; Koehler, R. T.; So, A. P.; Dube, S.; Rose, K. A.; Montesclaros, L.; Wang, S.; Stumbo, D. P.; Hodges, S. P.; et al. *Anal. Chem.* **2011**, *83*, 8604−8610.
(32) Jahn, M.; Seifert, J.; Hübschmann, T.; von Bergen, M.; Harms, H.; Müller, S. *Journal Of Integrated Omics* **2013**, *2*, 25−33.
(33) Koch, C.; Günther, S.; Desta, A. F.; Hübschmann, T.; Müller, S. *Nat. Protoc.* **2013**, *8*, 190−202.
(34) Hahne, F.; LeMeur, N.; Brinkman, R. R.; Ellis, B.; Haaland, P.; Sarkar, J.; Spidlen, J.; Strain, E.; Gentleman, R. *BMC Bioinformatics* **2009**, *10*, 106.
(35) Rozen, S.; Skaletsky, H. *Methods Mol. Biol.* **2000**, *132*, 365−386.
(36) Marshall, O. *Methods Mol. Biol.* **2007**, *402*, 403−414.
(37) Li, H.; Bo, H.; Wang, J.; Shao, H.; Huang, S. *Cytotechnology* **2011**, *63*, 7−12.
(38) Feldman, M. B.; Terry, D. S.; Altman, R. B.; Blanchard, S. C. *Nat. Chem. Biol.* **2010**, *6*, 54−62.
(39) Wilson, I. G. *Appl. Environ. Microbiol.* **1997**, *63*, 3741−3751.
(40) Weissensteiner, T.; Lanchbury, J. S. *BioTechniques* **1996**, *21*, 1102−1108.
(41) Bhat, S.; McLaughlin, J. L. H.; Emslie, K. R. *Analyst* **2011**, *136*, 724−732.
(42) Reichert-Schwillinsky, F.; Pin, C.; Dzieciol, M.; Wagner, M.; Hein, I. *Appl. Environ. Microbiol.* **2009**, *75*, 2132−2138.
(43) Volkmer, B.; Heinemann, M. *PLoS One* **2011**, *6*, e23126.
(44) Cormack, B. P.; Valdivia, R. H.; Falkow, S. *Gene* **1996**, *173*, 33−38.
(45) Iizuka, R.; Yamagishi-Shirasaki, M.; Funatsu, T. *Anal. Biochem.* **2011**, *414*, 173−178.
(46) Disney, M. D.; Seeberger, P. H. *Chemistry* **2004**, *10*, 3308−3314.
(47) Le Goffic, F.; Capmau, M. L.; Tangy, F.; Baillarge, M. *Eur. J. Biochem.* **1979**, *102*, 73−81.
(48) Misumi, M.; Nishimura, T.; Komai, T.; Tanaka, N. *Biochem. Biophys. Res. Commun.* **1978**, *84*, 358−365.
(49) Wong, C. H.; Hendrix, M.; Priestley, E. S.; Greenberg, W. A. *Chem. Biol.* **1998**, *5*, 397−406.
(50) Ling, J.; Cho, C.; Guo, L.; Aerni, H. R.; Rinehart, J.; Söll, D. *Mol. Cell* **2012**, *48*, 713−722.

3 Discussion

3.1 Enlarging the toolbox for microbial single cell analysis

Analysis of heterogeneity in microbial populations always depends on the toolbox that is available. Microbial populations in biotechnology have traditionally been characterized using bulk measurements such as western blotting to determine the amount of a target protein, or enzyme assays for the catalytic activity of cells. All these techniques are valuable and indispensable, but provide only an averaged view on the population. In the last decade the interest has shifted from the population to the single cell as the catalytic unit, and this development was boosted by methodical advances. The dominant techniques to analyze single cells are time-lapse (fluorescence) microscopy and flow cytometry. The major advantage of flow cytometry over microscopy is a higher throughput and the option to recover arbitrarily selected subpopulations *via* cell sorting (Müller & Nebe-von Caron, 2010).

3.1.1 Establishing mass spectrometric analysis of sorted subpopulations

In this work, the toolbox for single cell analysis of bacteria was significantly extended by combining cell sorting with downstream proteomics. The foundation of this work was laid by Wiacek *et al.* (2006) and Jehmlich *et al.* (2010), who were the first to sort bacterial cells for subsequent proteomics using 2D gel electrophoresis and mass spectrometry, respectively. The latter study already demonstrated higher recovery of proteins ($\sim$900 compared to $\sim$130) with significantly fewer cells (5×10^6 versus 1×10^9), but the sorted subpopulations were an artificial mixture of two species. Furthermore, pending questions were the long term stability of specimens stored for later analysis and the procedure of cell sampling/fixation in general. Countless protocols for cell sampling and storage are available, however, each of it can have a different impact on the sample quality.

A rapid method requiring no extra cooling is chemical fixation by ethanol, formaldehydes, inhibitors or metal ions (Günther *et al.*, 2008; Jensen *et al.*, 2010), while the preservation of intact and often viable cells is usually performed by deep freezing (De Paoli, 2005). A third way of sample preparation is (freeze) drying, which was shown to perform well for selected bacteria

and even preserve viability similar to deep freezing (Wong *et al.*, 2007; Bauer *et al.*, 2011). Three of these methods were applied to *Pseudomonas putida* (publication 1, page 34): Sodium azide fixation and storage at 4°C, deep freezing and storage at -80°C, and vacuum drying with storage at 4°C. Here, the goal was not to preserve cell viability but to keep cells in a state suitable for cytometry and mass spectrometry. This includes intact cells with identical scatter characteristics, DNA content and proteome profile. The result of this study was that deep freezing for short term (one week) and long term (one month) had only a minor effect on cell characteristics and protein composition, and was therefore rated best. Vacuum drying performed almost equally good, while sodium azide fixation was rather deleterious. This finding was in line with other studies who rated deep freezing using cryoprotectants as superior (Perlmutter *et al.*, 2004), even for mixed communities (Kerckhof *et al.*, 2014). However, chemical fixation of histological samples with ethanol was reported to be equally good for protein mass spectrometry (Chaurand *et al.*, 2008). This study significantly contributes to the knowledge of sample preparation in bacteria, and demonstrates once more that the impact of a sampling protocol is to be tested thoroughly beforehand. Furthermore, findings regarding *Pseudomonas putida* are most likely applicable to all other *Pseudomonas* strains and, probably, even to similar gram-negative bacteria. The optimized sampling protocol directly led to the first application of subpopulation proteomics to identify regulatory differences in different phenotypes of bacteria, here the producing and non-producing cells in a clonal population of *P. putida* KT2440 (discussed in section 3.2).

3.1.2 Digital PCR of sorted subpopulations

The second methodical advancement achieved in this work was the combination of cell sorting *via* flow cytometry and novel droplet digital PCR for precise quantification of DNA copy numbers. Digital PCR was already applied to quantify the copy number of very different DNA templates, among them also plasmids (Straub *et al.*, 2013). However, it nevertheless remains a young technology and, although it was shown that it is superior to conventional qRT-PCR, it is not trivial to assess the reliability and recovery of the method. How many copies of a template are really detected and what is the lowest number of template molecules? An elegant study by Bhat *et al.* (2011) used an artificial template with two independent target sequences to approach this question. The authors hypothesized, that 'molecular drop-out' would affect the two reactions individually, meaning that the failure of one reaction would still yield a positive signal of the other. And indeed, a PCR reaction under ideal conditions yielded only 88 to 94 % double positive signals. However, the portion of completely undetected template molecules remained unknown. These questions were answered by sorting a defined number of *P. putida* cells containing a genomic and a plasmid DNA template (publication 4, page 62). By precise deposition of exactly 1,000 cells with known DNA content, it was shown that only 75 % of template molecules from heat treated cells are detected. This is an important template inherent experimental error that needs

to be taken into account. Furthermore, a heat treatment protocol intended to crack bacterial cells (Carapuça *et al.*, 2007; Skulj *et al.*, 2008) was refined by optimizing heating duration for *P. putida* and *E. coli*. And finally, the power of digital PCR to detect very few template copies was challenged by depositing different numbers of cells ranging from 1 to 10,000. Theoretically, digital PCR can detect a single template molecule like the chromosome of a single bacterial cell in a background of 20,000 droplets (20 μL), but this study clearly demonstrated that false positive 'background' signals lead to extreme variability for 1 to 10 sorted cells. Quantification was reliable only from 100 cells upwards and optimal for 500-2,000 cells. As a matter of fact, sorting a single cell is technically feasible and reliable, but the detection of few or only one template molecules from that cell is not. These findings represent an important benchmark showing that detection of DNA from single sorted cells by PCR based methods will remain challenging.

3.2 Heterogeneity of *Pseudomonas putida* in a model bioprocess

3.2.1 Discrete subpopulations appear during protein production

The model bioprocess based on inducible *styA*/*styB* expression and implemented here in *P. putida* showed strong cell to cell variability from the beginning on. The formation of StyA after induction was tracked in live cells by a translational fusion with EGFP and always showed a variable portion of non-fluorescent cells (publication 2, page 43). More specifically, productivity was not normally distributed in the population but clustered in two very distinct subpopulations of high producers and non-producers. Such bimodal distributions of gene expression are a known phenomenon in biotechnological processes that can be caused by various forms of stress (Sevastsyanovich *et al.*, 2009), the induction regime (Wyre & Overton, 2014), noise on the regulatory level (Nikel *et al.*, 2014) or lack of genetic information (e.g. plasmid loss, Summers (1991)). The reason for bimodal distribution in this model bioprocess was completely unknown. However, the induction regime applied in this work was a standard concentration of a hydrophobic compound, sufficient to activate all cells in a culture. This may at least exclude the hypothesis of noisy gene expression. Another layer of heterogeneity is added by cell cycling. Staining of DNA yielded four subpopulations regarding DNA content which came up at different stages of cultivation. This is an amount of heterogeneity which can be expected, however, when StyA-EGFP productivity was correlated with DNA content, it was found that the most productive cells were not able to return to single chromosome content in stationary phase, but remained with a chromosome content of up to four copies. This effect could be due to defective cell division caused by inclusion bodies, as was already reported for *E. coli* (Lee *et al.*, 2008).

3.2.2 Protein production is related to various forms of stress

To gain a global insight and test as many probable causes for bimodal gene expression as possible, three subpopulations (EGFP negative, EPFG positive, EGFP positive with inclusion bodies) were sorted and analyzed by protein mass spectrometry. Altogether, the quantification of ~800 proteins was in line with comparable proteomic studies on *Pseudomonas* (Jehmlich *et al.*, 2010; Yun *et al.*, 2011). The surprising finding of this study was that almost no regulatory differences were found between subpopulations and only little changes in metabolic pathways were obvious. It can be concluded, that the morphologically very different subpopulations were not running different metabolic 'modes' or 'programs', but were functionally similar. Two groups of proteins were nevertheless extremely different between subpopulations. First and most strikingly, stress-related proteins were strongly up-regulated in producer cells, among them chaperones supposed to reduce protein stress, and catalases supposed to reduce oxidative stress. Thus, the top producers in a population react on heterologous gene expression with 'late-stage' counter measures, instead of redirecting carbon or energy flow in the first place. The second protein group gave a strong indication of the ultimate reason for bimodal gene expression. In the non-producing subpopulation, all four plasmid-encoded proteins were almost absent, pointing towards plasmid loss.

3.2.3 Variability in genome copy number is not represented on the protein level

Cell cycling is considered as a major cause for population heterogeneity (Müller & Nebe-von Caron, 2010). As discussed before (section 3.2), previous results showed a clear split of the *P. putida* KT2440 population into StyA-EGFP/StyB producing and non-producing cells. One hypothesis for lacking gene expression in a subpopulation was the occurrence of inactive cells seizing cell cycling and production. Related to this question, an experiment was –in collaboration with the University of Stuttgart– conceived to probe subpopulations with different DNA content, and to reveal underlying regulatory programs (publication 3, page 52). The wild type model strain *P. putida* KT2440 was grown under defined and stable conditions in chemostats and subpopulations were sorted regarding DNA content and further analyzed by mass spectrometry. Strikingly, almost no differences were found between subpopulations of one and the same growth rate. This allows two hypotheses: Either the cells are functionally identical although having a different number of chromosomes, and quickly shift between different cell cycle stages, while the proteome largely stays the same. This would be a very surprising finding, pointing towards great physiological homogeneity of all cells of the population. Or, 'true' single cell variability could not be resolved by sorting a subpopulation of millions of cells. The second hypothesis implies that single cell differences are averaged, or hidden subpopulations are overshadowed by a dominant

portion of cells. As the growth medium is constantly exchanged in a chemostat, non-dividing cells may be quickly washed out and displaced by the dividing subpopulation. However, a second finding of this study was that growth rate actually turned out to be the dominant factor shaping the proteome profile of the cells. Slow growing cells ($\mu = 0.1$) are in a completely different mode than fast growing cells ($\mu = 0.7$), the former desperately seeking nutrients, while the latter are channeling energy into replication.

3.2.4 Heterogeneity was ultimately caused by plasmid copy number fluctuation

Evidence from previous experiments pointed towards plasmid copy number (PCN) differences as the cause for population heterogeneity. A novel technique using droplet digital PCR was developed as outlined in section 3.1.2 to determine the absolute plasmid copy number in sorted subpopulations. After thorough optimization of the cell sorting protocol, cell number, DNA isolation, and ddPCR assay, the PCN in *P putida* KT2440 was finally determined in the course of target gene expression. The population was split into two subpopulations, and it could ultimately be shown that the plasmid was present in producers and absent in non-producers (publication 4, page 62).

What is the molecular reason for plasmid loss? In the first place, copy number and stability are linked with each other and determined by the replication system (Summers, 1991). Empirical evidence already suggested that the higher the copy number, the more stable is a plasmid during repetitive rounds of cell division, and the lower is the variability in PCN (Wong Ng *et al.*, 2010; Kittleson *et al.*, 2011). These insights are backed by the results of this study, as the PCN was, even in the most producing cells, surprisingly low (3-5). This was an overall unexpected finding for a plasmid series long used in biotechnology and reported to have medium to high copy number (Smits *et al.*, 2001). However, it is a shuttle vector containing two replication systems, one for *E. coli* and one for *Pseudomonas*, and it was shown that copy number was extremely high in *E. coli* (100-150), but higher as well in other Pseudomonas strains like DOT-T1E and VLB120 (10-20, unpublished results). Finally, these results demonstrate that PCN can vary a lot in different subpopulations, and averaged measurements will only provide a distorted view of the total population. Moreover, it is certainly not advisable to continue bioprocesses with this particular plasmid series in a host like *P putida* KT2440 as the very low copy number leads to segregational instability over time. The missing metabolic burden will confer a growth advantage to plasmid-free cells, whose fraction will exponentially increase over time (Lau *et al.*, 2013). However, the actual overall productivity was high despite the low copy number, showing that higher copy number is not automatically an advantage. Sufficient gene expression can be reached with low

copy numbers as well, accompanied by lower levels of stress (Jones *et al.*, 2000; Zdraljevic *et al.*, 2013).

3.3 Conclusion and future prospects

Cellular heterogeneity can significantly reduce efficiency and yield in industrial bioprocesses. In this project, population heterogeneity in *Pseudomonas putida* was investigated using an exemplary biotransformation from styrene to styrene oxide by production of an EGFP fused recombinant protein, StyA/StyB. Target gene expression was tracked *via* EGFP fluorescence and found to be extremely variable, showing discrete subpopulations of low and highly productive cells. Using cell sorting in combination with mass spectrometry or digital PCR, the phenotypic heterogeneity could be resolved on the molecular level providing insights into bottlenecks of recombinant protein production:

- The proteome profile of producing and non-producing cells was overall similar. Significant changes were, however, detected in stress-related proteins. Recombinant protein production was therefore a metabolic burden, which cells tried to overcome by launching corrective counter measures, rather than redirecting carbon or energy fluxes.

- Heterogeneity caused by cell cycling is a common phenomenon best seen by differential DNA content. However, subpopulations with one, two, or more than two copies of the chromosome did not show significant differences. This either underlines that all cycling cells of a population are essentially running the same 'program', or that true regulatory differences can only be detected by other means, such as single cell studies with reporter genes.

- The fundamental source of heterogeneity in this work was plasmid copy number variability. Non-producing cells appeared as a consequence of (segregational) plasmid loss, predominantly because the average plasmid copy number was much lower than originally expected.

Even if molecular causes for heterogeneity could be identified, many open questions remain. First of all, technical limitations allow only to analyze subpopulations with at least millions or hundreds of cells for mass spectrometry or digital PCR, respectively. This means, although single cells can be sorted with high precision, the result will nevertheless be an average of many cells. Furthermore, cell sorting requires the clear discrimination of subpopulations by certain morphological features. Therefore, getting quantitative information of many different molecules of *one and the same cell* remains challenging. Single cell studies using reporter genes and time lapse microscopy are one way, however, only very few proteins per cell at the same time can be tagged and monitored (Taniguchi *et al.*, 2010). Practically, this approach would be limited

to the observation of at most three important regulatory proteins. This could nevertheless be an alternative to detect cell cycle-related heterogeneity in a clonal population, which was only found to a small extent in this work using subpopulation mass spectrometry.

The efficiency of the model bioprocess studied here was spoiled by an actually trivial thing, an inappropriate plasmid expression system. Clearly, two ways can be seen to overcome such limitations in the biotechnology of the future. One is the ongoing trend away from variable and 'unstable' plasmids towards stable genomic integration or deletion of desired genes (Song *et al.*, 2015); A strategy which is more laborious in the beginning, but may yield the more reproducible process in the end. The molecular tool box for genome engineering is literally exploded in recent years. Traditional homologous recombination systems were supplemented by 'recombineering' using heterologous recombinases (e.g. λ Red), restriction enzyme mediated recombination (*I-Sce*1, TALENs, CRISPR-Cas), or RNA interference for gene knockdown (Martínez-García & de Lorenzo, 2011; Song *et al.*, 2015). Some of these techniques will probably become 'bread and butter' tools in molecular biology, boosting the power and diversity of microbial strains for biotechnology.

The other trend is towards using well-characterized and standardized biological 'parts'. As an example, the encountered bottleneck regarding plasmid copy number could be overcome by using a suite of rationally designed plasmids with known properties, such as the ones provided by the SEVA plasmid repository (Silva-Rocha *et al.*, 2013; Martínez-García *et al.*, 2015). Improved strains and better understanding of microbial physiology can finally pave the way for a more sustainable economy in the future.

4 References

Ackermann, M., Chao, L., Bergstrom, C. T. & Doebeli, M. *On the evolutionary origin of aging*. Aging Cell 6 (2): 235–44, **2007**.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection*. Mol Syst Biol 2: 2006.0008, **2006**.

Bauer, S. A. W., Schneider, S., Behr, J., Kulozik, U. & Foerst, P. *Combined influence of fermentation and drying conditions on survival and metabolic activity of starter and probiotic cultures after low-temperature vacuum drying*. J Biotechnol , **2011**.

Baumgarten, T., Sperling, S., Seifert, J., von Bergen, M., Steiniger, F., Wick, L. Y. & Heipieper, H. J. *Membrane vesicle formation as a multiple-stress response mechanism enhances Pseudomonas putida DOT-T1E cell surface hydrophobicity and biofilm formation*. Appl Environ Microbiol 78 (17): 6217–24, **2012**.

Bhat, S., McLaughlin, J. L. H. & Emslie, K. R. *Effect of sustained elevated temperature prior to amplification on template copy number estimation using digital polymerase chain reaction*. Analyst 136 (4): 724–32, **2011**.

Biobased-Industries-Consortium. *A Public-Private Partnership on Bio-based Industries*. Technical report, Biobased Industries Consortium, **2015**.

Blank, L. M., Ionidis, G., Ebert, B. E., Bühler, B. & Schmid, A. *Metabolic response of Pseudomonas putida during redox biocatalysis in the presence of a second octanol phase*. FEBS J 275 (20): 5173–90, **2008**.

Blattner, F. R., Plunkett, G. r., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. & Shao, Y. *The complete genome sequence of Escherichia coli K-12*. Science 277 (5331): 1453–62, **1997**.

Campos, M., Surovtsev, I. V, Kato, S., Paintdakhi, A., Beltran, B., Ebmeier, S. E. & Jacobs-Wagner, C. *A constant size extension drives bacterial cell size homeostasis*. Cell 159 (6): 1433–46, **2014**.

Carapuça, E., Azzoni, A. R., Prazeres, D. M. F., Monteiro, G. A. & Mergulhão, F. J. M. *Time-course determination of plasmid content in eukaryotic and prokaryotic cells using real-time PCR*. Mol Biotechnol 37 (2): 120–6, **2007**.

Cerrone, F., Duane, G., Casey, E., Davis, R., Belton, I., Kenny, S. T., Guzik, M. W., Woods, T., Babu, R. P. & O'Connor, K. *Fed-batch strategies using butyrate for high cell density cultivation of Pseudomonas putida and its use as a biocatalyst*. Appl Microbiol Biotechnol 98 (22): 9217–28, **2014**.

Chattopadhyay, P. K. & Roederer, M. *Cytometry: today's technology and tomorrow's horizons*. Methods 57 (3): 251–8, **2012**.

Chaurand, P., Latham, J. C., Lane, K. B., Mobley, J. A., Polosukhin, V. V., Wirth, P. S., Nanney, L. B. & Caprioli, R. M. *Imaging mass spectrometry of intact proteins from alcohol-preserved tissue specimens: bypassing formalin fixation*. J Proteome Res 7 (8): 3543–55, **2008**.

Chavarría, M., Kleijn, R. J., Sauer, U., Pflüger-Grau, K. & de Lorenzo, V. *Regulatory tasks of the phosphoenolpyruvate-phosphotransferase system of Pseudomonas putida in central carbon metabolism*. MBio 3 (2), **2012**.

Choi, H. J., Yoo, J.-S., Jeong, Y. K. & Joo, W. H. *Involvement of antioxidant defense system in solvent tolerance of Pseudomonas putida BCNU 106*. J Basic Microbiol 54 (9): 945–50, **2014**.

Chudakov, D. M., Matz, M. V., Lukyanov, S. & Lukyanov, K. A. *Fluorescent proteins and their applications in imaging living cells and tissues*. Physiol Rev 90 (3): 1103–63, **2010**.

Cooper, G. M. The Cell: A Molecular Approach. Sinauer Associates, 2nd edition, **2000**.

Cooper, S. & Helmstetter, C. E. *Chromosome replication and the division cycle of Escherichia coli B/r*. J Mol Biol 31 (3): 519–40, **1968**.

Cormack, B. P., Valdivia, R. H. & Falkow, S. *FACS-optimized mutants of the green fluorescent protein (GFP)*. Gene 173 (1 Spec No): 33–8, **1996**.

Darmon, E. & Leach, D. R. F. *Bacterial genome instability*. Microbiol Mol Biol Rev 78 (1): 1–39, **2014**.

Datsenko, K. A. & Wanner, B. L. *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products*. Proc Natl Acad Sci U S A 97 (12): 6640–5, **2000**.

De Paoli, P. *Bio-banking in microbiology: from sample collection to epidemiology, diagnosis and research*. FEMS Microbiol Rev 29 (5): 897–910, **2005**.

Drogue, B., Thomas, P., Balvay, L., Prigent-Combaret, C. & Dorel, C. *Engineering adherent bacteria by creating a single synthetic curli operon*. J Vis Exp 1 (69): e4176, **2012**.

Dusny, C., Fritzsch, F. S. O., Frick, O. & Schmid, A. *Isolated microbial single cells and resulting micropopulations grow faster in controlled environments*. Appl Environ Microbiol 78 (19): 7132–6, **2012**.

Díaz, L. F., Muñoz, R., Bordel, S. & Villaverde, S. *Toluene biodegradation by Pseudomonas putida F1: targeting culture stability in long-term operation*. Biodegradation 19 (2): 197–208, **2008**.

Ebert, B. E., Kurth, F., Grund, M., Blank, L. M. & Schmid, A. *Response of Pseudomonas putida KT2440 to increased NADH and ATP demand*. Appl Environ Microbiol 77 (18): 6597–605, **2011**.

Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. *Stochastic gene expression in a single cell*. Science 297 (5584): 1183–6, **2002**.

Eswaramoorthy, P., Winter, P. W., Wawrzusin, P., York, A. G., Shroff, H. & Ramamurthi, K. S. *Asymmetric division and differential gene expression during a bacterial developmental program requires DivIVA*. PLoS Genet 10 (8): e1004526, **2014**.

Exelus case study. *New Process for Producing Styrene Cuts Costs, Saves Energy, and Reduces Greenhouse Gas Emissions*. Technical report, U.S. Department of Energy, **2012**.

Friehs, K. *Plasmid copy number and plasmid stability*. Adv Biochem Eng Biotechnol 86: 47–82, **2004**.

Fritzsch, F. S. O., Rosenthal, K., Kampert, A., Howitz, S., Dusny, C., Blank, L. M. & Schmid, A. *Picoliter nDEP traps enable time-resolved contactless single bacterial cell analysis in controlled microenvironments*. Lab Chip 13 (3): 397–408, **2013**.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A. r., Smith, H. O. & Venter, J. C. *Creation of a bacterial cell controlled by a chemically synthesized genome*. Science 329 (5987): 52–6, **2010**.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. r. & Smith, H. O. *Enzymatic assembly of DNA molecules up to several hundred kilobases*. Nat Methods 6 (5): 343–5, **2009**.

Greated, A., Lambertsen, L., Williams, P. A. & Thomas, C. M. *Complete sequence of the IncP-9 TOL plasmid pWW0 from Pseudomonas putida*. Environ Microbiol 4 (12): 856–71, **2002**.

Günther, S., Hübschmann, T., Rudolf, M., Eschenhagen, M., Röske, I., Harms, H. & Müller, S. *Fixation procedures for flow cytometric analysis of environmental bacteria*. J Microbiol Methods 75 (1): 127–34, **2008**.

Hedhammar, M., Stenvall, M., Lönneborg, R., Nord, O., Sjölin, O., Brismar, H., Uhlén, M., Ottosson, J. & Hober, S. *A novel flow cytometry-based method for analysis of expression levels in Escherichia coli, giving information about precipitated and soluble protein*. J Biotechnol 119 (2): 133–46, **2005**.

Heipieper, H. J., Meinhardt, F. & Segura, A. *The cis-trans isomerase of unsaturated fatty acids in Pseudomonas and Vibrio: biochemistry, molecular biology and physiological function of a unique stress adaptive mechanism*. FEMS Microbiol Lett 229 (1): 1–7, **2003**.

Henry, J. T. & Crosson, S. *Chromosome replication and segregation govern the biogenesis and inheritance of inorganic polyphosphate granules*. Mol Biol Cell 24 (20): 3177–86, **2013**.

Hewitt, C. J., Onyeaka, H., Lewis, G., Taylor, I. W. & Nienow, A. W. *A comparison of high cell density fed-batch fermentations involving both induced and non-induced recombinant Escherichia coli under well-mixed small-scale and simulated poorly mixed large-scale conditions*. Biotechnol Bioeng 96 (3): 495–505, **2007**.

Hilfinger, A. & Paulsson, J. *Separating intrinsic from extrinsic fluctuations in dynamic biological systems*. Proc Natl Acad Sci U S A 108 (29): 12167–72, **2011**.

Hindson, B. J., Ness, K. D., Masquelier, D. A., Belgrader, P., Heredia, N. J., Makarewicz, A. J., Bright, I. J., Lucero, M. Y., Hiddessen, A. L., Legler, T. C., Kitano, T. K., Hodel, M. R., Petersen, J. F., Wyatt, P. W., Steenblock, E. R., Shah, P. H., Bousse, L. J., Troup, C. B., Mellen, J. C., Wittmann, D. K., Erndt, N. G., Cauley, T. H., Koehler, R. T., So, A. P., Dube, S., Rose, K. A., Montesclaros, L., Wang, S., Stumbo, D. P., Hodges, S. P., Romine, S., Milanovich, F. P., White, H. E., Regan, J. F., Karlin-Neumann, G. A., Hindson, C. M., Saxonov, S. & Colston, B. W. *High-throughput droplet digital PCR system for absolute quantitation of DNA copy number*. Anal Chem 83 (22): 8604–10, **2011**.

Hindson, C. M., Chevillet, J. R., Briggs, H. A., Gallichotte, E. N., Ruf, I. K., Hindson, B. J., Vessella, R. L. & Tewari, M. *Absolute quantification by droplet digital PCR versus analog real-time PCR*. Nat Methods , **2013**.

Huh, D. & Paulsson, J. *Non-genetic heterogeneity from stochastic partitioning at cell division*. Nat Genet 43 (2): 95–100, **2011**a.

Huh, D. & Paulsson, J. *Random partitioning of molecules at cell division*. Proc Natl Acad Sci U S A 108 (36): 15004–9, **2011**b.

IGEM. *Engineering banana and wintergreen scent in Escherichia coli*. http://openwetware.org/wiki/IGEM:MIT/2006/Blurb. **2006**.

Jehmlich, N., Hübschmann, T., Gesell Salazar, M., Völker, U., Benndorf, D., Müller, S., von Bergen, M. & Schmidt, F. *Advanced tool for characterization of microbial cultures by combining cytomics and proteomics*. Appl Microbiol Biotechnol 88 (2): 575–84, **2010**.

Jensen, U. B., Owens, D. M., Pedersen, S. & Christensen, R. *Zinc fixation preserves flow cytometry scatter and fluorescence parameters and allows simultaneous analysis of DNA content and synthesis, and intracellular and surface epitopes*. Cytometry A 77 (8): 798–804, **2010**.

Jones, K. L., Kim, S. W. & Keasling, J. D. *Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria*. Metab Eng 2 (4): 328–38, **2000**.

Juurik, T., Ilves, H., Teras, R., Ilmjärv, T., Tavita, K., Ukkivi, K., Teppo, A., Mikkel, K. & Kivisaar, M. *Mutation frequency and spectrum of mutations vary at different chromosomal positions of Pseudomonas putida*. PLoS One 7 (10): e48511, **2012**.

Kerckhof, F.-M., Courtens, E. N. P., Geirnaert, A., Hoefman, S., Ho, A., Vilchez-Vargas, R., Pieper, D. H., Jauregui, R., Vlaeminck, S. E., Van de Wiele, T., Vandamme, P., Heylen, K. & Boon, N. *Optimized cryopreservation of mixed microbial communities for conserved functionality and diversity*. PLoS One 9 (6): e99517, **2014**.

Kim, B., Park, H., Na, D. & Lee, S. Y. *Metabolic engineering of Escherichia coli for the production of phenol from glucose*. Biotechnol J 9 (5): 621–9, **2014**.

Kittleson, J. T., Cheung, S. & Anderson, J. C. *Rapid optimization of gene dosage in E. coli using DIAL strains*. J Biol Eng 5: 10, **2011**.

Knight, T. F. *Idempotent Vector Design for Standard Assembly of BioBricks*. Technical report, MIT Synthetic Biology Working Group, **2003**.

Kroll, J., Klinter, S., Schneider, C., Voss, I. & Steinbüchel, A. *Plasmid addiction systems: perspectives and applications in biotechnology*. Microb Biotechnol 3 (6): 634–57, **2010**.

Lau, B. T. C., Malkus, P. & Paulsson, J. *New quantitative methods for measuring plasmid loss rates reveal unexpected stability*. Plasmid 70 (3): 353–61, **2013**.

Lee, H., Popodi, E., Tang, H. & Foster, P. L. *Rate and molecular spectrum of spontaneous mutations in the bacterium Escherichia coli as determined by whole-genome sequencing*. Proc Natl Acad Sci U S A 109 (41): E2774–83, **2012**.

Lee, K. K., Jang, C. S., Yoon, J. Y., Kim, S. Y., Kim, T. H., Ryu, K. H. & Kim, W. *Abnormal cell division caused by inclusion bodies in E. coli; increased resistance against external stress*. Microbiol Res 163 (4): 394–402, **2008**.

Lenz, P. & Søgaard-Andersen, L. *Temporal and spatial oscillations in bacteria*. Nat Rev Microbiol 9 (8): 565–77, **2011**.

Lin, B., Lyu, J., Lyu, X.-j., Yu, H.-q., Hu, Z., Lam, J. C. W. & Lam, P. K. S. *Characterization of cefalexin degradation capabilities of two Pseudomonas strains isolated from activated sludge*. J Hazard Mater 282: 158–64, **2015**.

de Lorenzo, V., Herrero, M., Sanchez, J. M. & Timmis, K. N. *Mini-transposons in microbial ecology and environmental biotechnology*. FEMS Microbiology Ecology 27 (3): 211–224, **1998**.

Makino, T., Skretas, G. & Georgiou, G. *Strain engineering for improved expression of recombinant proteins in bacteria*. Microb Cell Fact 10: 32, **2011**.

Mandalakis, M., Panikov, N., Dai, S., Ray, S. & Karger, B. L. *Comparative proteomic analysis reveals mechanistic insights into Pseudomonas putida F1 growth on benzoate and citrate*. AMB Express 3 (1): 64, **2013**.

Martínez-García, E., Aparicio, T., Goñi-Moreno, A., Fraile, S. & de Lorenzo, V. *SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities*. Nucleic Acids Res 43 (Database issue): D1183–9, **2015**.

Martínez-García, E. & de Lorenzo, V. *Engineering multiple genomic deletions in Gram-negative bacteria: analysis of the multi-resistant antibiotic profile of Pseudomonas putida KT2440*. Environ Microbiol 13 (10): 2702–16, **2011**.

Martínez-García, E. & de Lorenzo, V. *Transposon-based and plasmid-based genetic tools for editing genomes of gram-negative bacteria*. Methods Mol Biol 813: 267–83, **2012**.

Meijnen, J.-P., de Winde, J. H. & Ruijssenaars, H. J. *Engineering Pseudomonas putida S12 for efficient utilization of D-xylose and L-arabinose*. Appl Environ Microbiol 74 (16): 5031–7, **2008**.

Meijnen, J.-P., de Winde, J. H. & Ruijssenaars, H. J. *Establishment of oxidative D-xylose metabolism in Pseudomonas putida S12*. Appl Environ Microbiol 75 (9): 2784–91, **2009**.

Mi, J., Becher, D., Lubuta, P., Dany, S., Tusch, K., Schewe, H., Buchhaupt, M. & Schrader, J. *De novo production of the monoterpenoid geranic acid by metabolically engineered Pseudomonas putida*. Microb Cell Fact 13 (1): 170, **2014**.

Molina, L., Duque, E., Gómez, M. J., Krell, T., Lacal, J., García-Puente, A., García, V., Matilla, M. A., Ramos, J.-L. & Segura, A. *The pGRT1 plasmid of Pseudomonas putida DOT-T1E encodes functions relevant for survival under harsh conditions in the environment*. Environ Microbiol 13 (8): 2315–27, **2011**.

Morris, J. J., Papoulis, S. E. & Lenski, R. E. *Coexistence of evolving bacteria stabilized by a shared black queen function*. Evolution 68 (10): 2960–71, **2014**.

Müller, S. & Nebe-von Caron, G. *Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities*. FEMS Microbiol Rev 34 (4): 554–87, **2010**.

Müller, S., Harms, H. & Bley, T. *Origin and analysis of microbial population heterogeneity in bioprocesses*. Curr Opin Biotechnol 21 (1): 100–13, **2010**.

Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V. A. P., Fouts, D. E., Gill, S. R., Pop, M., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R. T., 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, J. A., Timmis, K. N., Düsterhöft, A., Tümmler, B. & Fraser, C. M. *Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440*. Environ Microbiol 4 (12): 799–808, **2002**.

Nikel, P. I., Silva-Rocha, R., Benedetti, I. & de Lorenzo, V. *The private life of environmental bacteria: pollutant biodegradation at the single cell level*. Environ Microbiol 16 (3): 628–42, **2014**.

Nikodinovic-Runic, J., Flanagan, M., Hume, A. R., Cagney, G. & O'Connor, K. E. *Analysis of the Pseudomonas putida CA-3 proteome during growth on styrene under nitrogen-limiting and non-limiting conditions*. Microbiology 155 (Pt 10): 3348–61, **2009**.

Otto, K., Hofstetter, K., Röthlisberger, M., Witholt, B. & Schmid, A. *Biochemical characterization of StyAB from Pseudomonas sp. strain VLB120 as a two-component flavin-diffusible monooxygenase*. J Bacteriol 186 (16): 5292–302, **2004**.

Panke, S., de Lorenzo, V., Kaiser, A., Witholt, B. & Wubbolts, M. G. *Engineering of a stable whole-cell biocatalyst capable of (S)-styrene oxide formation for continuous two-liquid-phase applications*. Appl Environ Microbiol 65 (12): 5619–23, **1999**.

Panke, S., Witholt, B., Schmid, A. & Wubbolts, M. G. *Towards a biocatalyst for (S)-styrene oxide production: characterization of the styrene degradation pathway of Pseudomonas sp. strain VLB120*. Appl Environ Microbiol 64 (6): 2032–43, **1998**.

Panke, S., Wubbolts, M. G., Schmid, A. & Witholt, B. *Production of enantiopure styrene oxide by recombinant Escherichia coli synthesizing a two-component styrene monooxygenase*. Biotechnol Bioeng 69 (1): 91–100, **2000**.

Park, J.-B., Bühler, B., Habicher, T., Hauer, B., Panke, S., Witholt, B. & Schmid, A. *The efficiency of recombinant Escherichia coli as biocatalyst for stereospecific epoxidation*. Biotechnol Bioeng 95 (3): 501–12, **2006**.

Perlmutter, M. A., Best, C. J. M., Gillespie, J. W., Gathright, Y., González, S., Velasco, A., Linehan, W. M., Emmert-Buck, M. R. & Chuaqui, R. F. *Comparison of snap freezing versus ethanol fixation for gene expression profiling of tissue specimens*. J Mol Diagn 6 (4): 371–7, **2004**.

Petrosino, J. F., Galhardo, R. S., Morales, L. D. & Rosenberg, S. M. *Stress-induced beta-lactam antibiotic resistance mutation and sequences of stationary-phase mutations in the Escherichia coli chromosome*. J Bacteriol 191 (19): 5881–9, **2009**.

Poblete-Castro, I., Becker, J., Dohnt, K., dos Santos, V. M. & Wittmann, C. *Industrial biotechnology of Pseudomonas putida and related species*. Appl Microbiol Biotechnol 93 (6): 2279–90, **2012**.

Poblete-Castro, I., Binger, D., Oehlert, R. & Rohde, M. *Comparison of mcl-Poly(3-hydroxyalkanoates) synthesis by different Pseudomonas putida strains from crude glycerol: citrate accumulates at high titer under PHA-producing conditions*. BMC Biotechnol 14 (1): 110, **2014**.

Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. *Primary structure of the Aequorea victoria green-fluorescent protein*. Gene 111 (2): 229–33, **1992**.

Quiñones-Valles, C., Sánchez-Osorio, I. & Martínez-Antonio, A. *Dynamical modeling of the cell cycle and cell fate emergence in Caulobacter crescentus*. PLoS One 9 (11): e111116, **2014**.

Raeside, C., Gaffé, J., Deatherage, D. E., Tenaillon, O., Briska, A. M., Ptashkin, R. N., Cruveiller, S., Médigue, C., Lenski, R. E., Barrick, J. E. & Schneider, D. *Large chromosomal rearrangements during a long-term evolution experiment with Escherichia coli*. MBio 5 (5): e01377–14, **2014**.

Rando, O. J. & Verstrepen, K. J. *Timescales of genetic and epigenetic inheritance*. Cell 128 (4): 655–68, **2007**.

Ranquet, C., Toussaint, A., de Jong, H., Maenhaut-Michel, G. & Geiselmann, J. *Control of bacteriophage mu lysogenic repression*. J Mol Biol 353 (1): 186–95, **2005**.

Reinhard, F. & van der Meer, J. R. *Life history analysis of integrative and conjugative element activation in growing microcolonies of Pseudomonas*. J Bacteriol 196 (7): 1425–34, **2014**.

Rühl, J., Schmid, A. & Blank, L. M. *Selected Pseudomonas putida strains able to grow in the presence of high butanol concentrations*. Appl Environ Microbiol 75 (13): 4653–6, **2009**.

Segura, A., Molina, L., Fillet, S., Krell, T., Bernal, P., Muñoz-Rojas, J. & Ramos, J.-L. *Solvent tolerance in Gram-negative bacteria*. Curr Opin Biotechnol 23 (3): 415–21, **2012**.

Serrano, L. *Synthetic biology: promises and challenges*. Mol Syst Biol 3: 158, **2007**.

Sevastsyanovich, Y., Alfasi, S., Overton, T., Hall, R., Jones, J., Hewitt, C. & Cole, J. *Exploitation of GFP fusion proteins and stress avoidance as a generic strategy for the production of high-quality recombinant proteins*. FEMS Microbiol Lett 299 (1): 86–94, **2009**.

Shapiro, H. M. Practical Flow Cytometry. John Wiley & Sons, 4th edition, **2003**.

Shields, C. W. t., Reyes, C. D. & López, G. P. *Microfluidic cell sorting: a review of the advances in the separation of cells from debulking to rare cell isolation*. Lab Chip 15 (5): 1230–49, **2015**.

Shimomura, O., JOHNSON, F. H. & SAIGA, Y. *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea victoria*. J Cell Comp Physiol 59: 223–39, **1962**.

Siedler, S., Schendzielorz, G., Binder, S., Eggeling, L., Bringer, S. & Bott, M. *SoxR as a single-cell biosensor for NADPH-consuming enzymes in Escherichia coli*. ACS Synth Biol 3 (1): 41–7, **2014**.

Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de Las Heras, A., Páez-Espino, A. D., Durante-Rodríguez, G., Kim, J., Nikel, P. I., Platero, R. & de Lorenzo, V. *The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes*. Nucleic Acids Res 41 (Database issue): D666–75, **2013**.

Simon, O., Klaiber, I., Huber, A. & Pfannstiel, J. *Comprehensive proteome analysis of the response of Pseudomonas putida KT2440 to the flavor compound vanillin*. J Proteomics 109: 212–27, **2014**.

Skarstad, K., Steen, H. B. & Boye, E. *Escherichia coli DNA distributions measured by flow cytometry and compared with theoretical computer simulations*. J Bacteriol 163 (2): 661–8, **1985**.

Skulj, M., Okrslar, V., Jalen, S., Jevsevar, S., Slanc, P., Strukelj, B. & Menart, V. *Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes*. Microb Cell Fact 7: 6, **2008**.

Smits, T. H., Röthlisberger, M., Witholt, B. & van Beilen, J. B. *Molecular screening for alkane hydroxylase genes in Gram-negative and Gram-positive strains*. Environ Microbiol 1 (4): 307–17, **1999**.

Smits, T. H., Seeger, M. A., Witholt, B. & van Beilen, J. B. *New alkane-responsive expression vectors for Escherichia coli and pseudomonas*. Plasmid 46 (1): 16–24, **2001**.

Song, C. W., Lee, J. & Lee, S. Y. *Genome engineering and gene expression control for bacterial strain development*. Biotechnol J 10 (1): 56–68, **2015**.

Stewart, E. J., Madden, R., Paul, G. & Taddei, F. *Aging and death in an organism that reproduces by morphologically symmetric division*. PLoS Biol 3 (2): e45, **2005**.

Strain, M. C., Lada, S. M., Luong, T., Rought, S. E., Gianella, S., Terry, V. H., Spina, C. A., Woelk, C. H. & Richman, D. D. *Highly precise measurement of HIV DNA by droplet digital PCR*. PLoS One 8 (4): e55943, **2013**.

Straub, T., Baird, C., Bartholomew, R. A., Colburn, H., Seiner, D., Victry, K., Zhang, L. & Bruckner-Lea, C. J. *Estimated copy number of Bacillus anthracis plasmids pXO1 and pXO2 using digital PCR*. J Microbiol Methods 92 (1): 9–10, **2013**.

Strovas, T. J. & Lidstrom, M. E. *Population heterogeneity in Methylobacterium extorquens AM1*. Microbiology 155 (Pt 6): 2040–8, **2009**.

Summers, D. K. *The kinetics of plasmid loss*. Trends Biotechnol 9 (8): 273–8, **1991**.

Surmann, K., Michalik, S., Hildebrandt, P., Gierok, P., Depke, M., Brinkmann, L., Bernhardt, J., Salazar, M. G., Sun, Z., Shteynberg, D., Kusebauch, U., Moritz, R. L., Wollscheid, B., Lalk, M., Völker, U. & Schmidt, F. *Comparative proteome analysis reveals conserved and specific adaptation patterns of Staphylococcus aureus after internalization by different types of human non-professional phagocytic host cells*. Front Microbiol 5: 392, **2014**.

Tan, I. S. & Ramamurthi, K. S. *Spore formation in Bacillus subtilis*. Environ Microbiol Rep 6 (3): 212–25, **2014**.

Taniguchi, Y., Choi, P. J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A. & Xie, X. S. *Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells*. Science 329 (5991): 533–8, **2010**.

Tao, F., Shen, Y., Fan, Z., Tang, H. & Xu, P. *Genome sequence of Pseudomonas putida S12, a potential platform strain for industrial production of valuable chemicals*. J Bacteriol 194 (21): 5985–6, **2012**.

Tsokos, C. G. & Laub, M. T. *Polarity and cell fate asymmetry in Caulobacter crescentus*. Curr Opin Microbiol 15 (6): 744–50, **2012**.

Tsoy, O., Yurieva, M., Kucharavy, A., O'Reilly, M. & Mushegian, A. *Minimal genome encoding proteins with constrained amino acid repertoire*. Nucleic Acids Res 41 (18): 8444–51, **2013**.

Udaondo, Z., Duque, E., Fernández, M., Molina, L., de la Torre, J., Bernal, P., Niqui, J.-L., Pini, C., Roca, A., Matilla, M. A., Molina-Henares, M. A., Silva-Jiménez, H., Navarro-Avilés, G., Busch, A., Lacal, J., Krell, T., Segura, A. & Ramos, J.-L. *Analysis of solvent tolerance in Pseudomonas putida DOT-T1E based on its genome sequence and a collection of mutants*. FEBS Lett 586 (18): 2932–8, **2012**.

Veening, J.-W., Smits, W. K. & Kuipers, O. P. *Bistability, epigenetics, and bet-hedging in bacteria*. Annu Rev Microbiol 62: 193–210, **2008**.

Vogelstein, B. & Kinzler, K. W. *Digital PCR*. Proc Natl Acad Sci U S A 96 (16): 9236–41, **1999**.

Volmer, J., Neumann, C., Bühler, B. & Schmid, A. *Engineering of Pseudomonas taiwanensis VLB120 for constitutive solvent tolerance and increased specific styrene epoxidation activity*. Appl Environ Microbiol 80 (20): 6539–48, **2014**.

Wakamoto, Y., Dhar, N., Chait, R., Schneider, K., Signorino-Gelo, F., Leibler, S. & McKinney, J. D. *Dynamic persistence of antibiotic-stressed mycobacteria*. Science 339 (6115): 91–5, **2013**.

Wang, J. D. & Levin, P. A. *Metabolism, cell growth and the bacterial cell cycle*. Nat Rev Microbiol 7 (11): 822–7, **2009**.

Want, A., Hancocks, H., Thomas, C. R., Stocks, S. M., Nebe-von Caron, G. & Hewitt, C. J. *Multi-parameter flow cytometry and cell sorting reveal extensive physiological heterogeneity in Bacillus cereus batch cultures*. Biotechnol Lett 33 (7): 1395–405, **2011**.

Whale, A. S., Huggett, J. F., Cowen, S., Speirs, V., Shaw, J., Ellison, S., Foy, C. A. & Scott, D. J. *Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation*. Nucleic Acids Res 40 (11): e82, **2012**.

Wiacek, C., Müller, S. & Benndorf, D. *A cytomic approach reveals population heterogeneity of Cupriavidus necator in response to harmful phenol concentrations*. Proteomics 6 (22): 5983–94, **2006**.

Wielgoss, S., Barrick, J. E., Tenaillon, O., Wiser, M. J., Dittmar, W. J., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R. E. & Schneider, D. *Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load*. Proc Natl Acad Sci U S A 110 (1): 222–7, **2013**.

Wijte, D., van Baar, B. L. M., Heck, A. J. R. & Altelaar, A. F. M. *Probing the proteome response to toluene exposure in the solvent tolerant Pseudomonas putida S12*. J Proteome Res 10 (2): 394–403, **2011**.

Williams, P. A. & Murray, K. *Metabolism of benzoate and the methylbenzoates by Pseudomonas putida (arvilla) mt-2: evidence for the existence of a TOL plasmid*. J Bacteriol 120 (1): 416–23, **1974**.

Winsor, G. L., Lam, D. K. W., Fleming, L., Lo, R., Whiteside, M. D., Yu, N. Y., Hancock, R. E. W. & Brinkman, F. S. L. *Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes*. Nucleic Acids Res 39 (Database issue): D596–600, **2011**.

Wong, Y.-L., Sampson, S., Germishuizen, W. A., Goonesekera, S., Caponetti, G., Sadoff, J., Bloom, B. R. & Edwards, D. *Drying a tuberculosis vaccine without freezing*. Proc Natl Acad Sci U S A 104 (8): 2591–5, **2007**.

Wong Ng, J., Chatenay, D., Robert, J. & Poirier, M. G. *Plasmid copy number noise in monoclonal populations of bacteria*. Phys Rev E Stat Nonlin Soft Matter Phys 81 (1 Pt 1): 011909, **2010**.

Wyre, C. & Overton, T. W. *Use of a stress-minimisation paradigm in high cell density fed-batch Escherichia coli fermentations to optimise recombinant protein production*. J Ind Microbiol Biotechnol 41 (9): 1391–404, **2014**.

Yun, S.-H., Park, G. W., Kim, J. Y., Kwon, S. O., Choi, C.-W., Leem, S.-H., Kwon, K.-H., Yoo, J. S., Lee, C., Kim, S. & Kim, S. I. *Proteomic characterization of the Pseudomonas putida KT2440 global response to a monocyclic aromatic compound by iTRAQ analysis and 1DE-MudPIT*. J Proteomics 74 (5): 620–8, **2011**.

Zdraljevic, S., Wagner, D., Cheng, K., Ruohonen, L., Jäntti, J., Penttilä, M., Resnekov, O. & Pesce, C. G. *Single-cell measurements of enzyme levels as a predictive tool for cellular fates during organic acid production*. Appl Environ Microbiol 79 (24): 7569–82, **2013**.

Zhang, H., Lin, M., Shi, H., Ji, W., Huang, L., Zhang, X., Shen, S., Gao, R., Wu, S., Tian, C., Yang, Z., Zhang, G., He, S., Wang, H., Saw, T., Chen, Y. & Ouyang, Q. *Programming a Pavlovian-like conditioning circuit in Escherichia coli*. Nat Commun 5: 3102, **2014**.

5 Appendix

Declaration of authorship

I herewith declare that

- I have written this thesis autonomously incorporating my own ideas and judgments; I have made use of no other resources than stated and direct or indirect quotations from other work have been marked accordingly; full reference of their source has been provided in the proper way.

- all persons are listed that provided me with support for the selection and evaluation of the material for my thesis; nature and scope of my own contribution and the share of the co-authors is listed in 'Author contributions of published articles' (5.1).

- no other persons have provided support and thereby contributed to the thesis; in particular, no PhD consultants were used, and no third party has received direct or indirect financial benefits in goods and services for work that stands in relation to the work presented in the thesis.

- this thesis has not been submitted in an equal or similar form for examination for the degree of doctorate or any other degree at another academic institution, and has not been published.

- no further unsuccessful doctoral examination process has taken place.

Place   Date   Signature
Ort   Datum   Unterschrift

5.1 Author contributions of published articles

Author contributions for the four in this work enclosed publications are listed on the following pages:
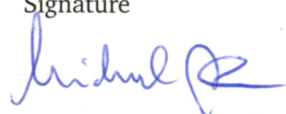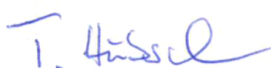
- Publication 1, page 89.

- Publication 2, page 90.

- Publication 3, page 91.

- Publication 4, page 92.

Nachweis über Anteile der Koautoren, Michael Jahn

Characterization of population heterogeneity in a model biotechnological process using *Pseudomonas putida*

Documentation of contingents of coauthors

Title: **Comparison of preservation methods for bacterial cells in cytomics and proteomics**

Journal: Journal Of Integrated Omics

Authors: Michael Jahn, Jana Seifert, Thomas Hübschmann, Martin von Bergen, Hauke Harms, Susann Müller
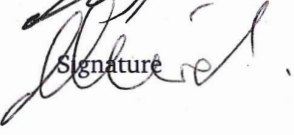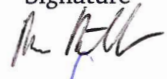
Michael Jahn	Study concept Cultivation, sampling Flow cytometry, cell sorting Data analysis Manuscript writing	Signature
Jana Seifert	Study concept Proteomics Data analysis Manuscript revision	Signature
Thomas Hübschmann	Flow cytometry, cell sorting Manuscript revision	
Martin von Bergen	Study concept Manuscript revision	Signature
Hauke Harms	Study concept Manuscript revision	Signature
Susann Müller	Study concept Manuscript writing	Signature

Nachweis über Anteile der Koautoren, Michael Jahn
Characterization of population heterogeneity in a model biotechnological process using *Pseudomonas putida*

Documentation of contingents of coauthors

Title:	**Subpopulation-proteomics in prokaryotic populations**
Journal:	Current Opinion in Biotechnology
Authors:	Michael Jahn, Jana Seifert, Martin von Bergen, Andreas Schmid, Bruno Bühler, Susann Müller

Michael Jahn
Study concept
Cultivation, sampling
Flow cytometry, cell sorting
Data analysis
Manuscript writing

Signature

Jana Seifert
Study concept
Proteomics
Data analysis
Manuscript writing

Signature

Martin von Bergen
Manuscript writing

Signature

Andreas Schmid
Manuscript writing

Signature

Bruno Bühler
Manuscript writing

Signature

Susann Müller
Study concept
Manuscript writing

Signature

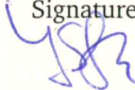**Nachweis über Anteile der Koautoren, Michael Jahn**
Characterization of population heterogeneity in a model biotechnological process using *Pseudomonas putida*

Documentation of contingents of coauthors

Title: **Subpopulation-proteomics reveal growth rate, but not cell cycling, as a major impact on protein composition in *Pseudomonas putida* KT2440**

Journal: AMB Express

Authors: Sarah Lieder, Michael Jahn, Jana Seifert, Martin von Bergen, Susann Müller, Ralf Takors

Sarah Lieder	Study concept Cultivation, sampling Determination of exp. variables Data analysis Manuscript writing	Signature
Michael Jahn	Study concept Flow cytometry, cell sorting Data analysis Manuscript writing	Signature
Jana Seifert	Proteomics Data analysis Manuscript revision	Signature
Martin von Bergen	Manuscript revision	Signature
Susann Müller	Study concept Manuscript writing	Signature
Ralf Takors	Study concept Manuscript writing	Signature

Nachweis über Anteile der Koautoren, Michael Jahn
Characterization of population heterogeneity in a model biotechnological process using *Pseudomonas putida*

Documentation of contingents of coauthors

Title:	**Accurate determination of plasmid copy number of f ow-sorted cells using droplet digital PCR**
Journal:	Analytical Chemistry
Authors:	Michael Jahn, Carsten Vorpahl, Dominique Türkowsky, Martin Lindmeyer, Bruno Bühler, Hauke Harms, Susann Müller

Michael Jahn	Study concept Cultivation, sampling Flow cytometry, cell sorting Digital PCR Data analysis Manuscript writing	Signature
Carsten Vorpahl	Digital PCR, qRT-PCR Data analysis Manuscript revision	Signature
Dominique Türkowsky	Digital PCR, qRT-PCR Data analysis Manuscript revision	Signature
Martin Lindmeyer	Plasmid construction Manuscript revision	Signature
Bruno Bühler	Manuscript revision	Signature
Hauke Harms	Manuscript revision	Signature
Susann Müller	Study concept Manuscript writing	Signature

5.2 Curriculum vitae

Personal data

Name:	Michael Jahn
Date of birth:	Dec 31st, 1985
Born in:	Dresden, Germany
Current residence:	Oststraße 96, 04317 Leipzig
Phone:	+49 341 26457111
Mobile phone:	+49 173 6919919
Email:	michael.jahn@ufz.de
Private email:	michael.jahn.dresden@gmail.com

Education

2011–2015	PhD student in the group of Flow Cytometry, Dept. Environmental Microbiology, Helmholtz-Centre for Environmental Research
2005–2011	Diploma studies, Biology, Dresden University of Technology (TUD) Main subjects: genetics, biochemistry, immunology. Diploma thesis: 'Diffusion of α-factor mating pheromone in a two component yeast biosensor system' Grade: 1.2
1996–2004	Gymnasium Romain-Rolland, Dresden, 05/2004 Abitur (grade 1.6)
1992–1996	Elementary school, Dresden

Work experience

Since 2011	Supervision of bachelor and master students
08-12/2009	Erasmus placement: Center of Excellence in Evolutionary Genetics and Physiology, Prof. Nikinmaa, Dept. of Biology, University of Turku, Finland
2008–2009	Student assistent: Institute of Genetics, Dept. of Biology at TUD
2007–2008	Student assistent: Mitteldeutscher Praxisverbund Humangenetik, Dresden, human genetics lab
2004–2005	Voluntary service: National park 'Sächsische Schweiz', tour guide and education

Languages

German	Native language
English	Proficient user, CEF level: C1
French	Basic user, CEF level: A1-A2
Polish	Basic user, CEF level: A1

Qualifications & Skills

Courses and Trainings	Laser scanning microscopy, 1 week intensive course, 2013 Approved project leader for genetic works, according to German laws (§ 14,15 GenTSV), 2015
Molecular genetics	Cultivation of microbes (*E. coli*, *P. putida*, *S. cerevisiae*) Cultivation of human cancer cells Molecular genetics: DNA, RNA, protein extraction and purification Vector design, DNA cloning, recombinant gene expression Droplet Digital PCR, qRT-PCR, gel electrophoresis Fluorescence microscopy
High-throughput techniques	Flow cytometry, cell sorting Mass spectrometry, proteomics Metabolic mapping, data mining, visualization Automated microscopic imaging and data mining
Software	Programming: R (adv.), Latex (adv.), Python (beg.) Office: Microsoft Office, OpenOffice Imaging: CellProfiler, ImageJ, Inkscape, GIMP Flow cytometry: R flow packages, Summit, FlowJo Genome and proteome databases: NCBI, DAVID, KEGG, BioCyc

5.3 List of publications and conference contributions

Publications

Jahn, Michael; Mölle, Annett; Rödel, Gerhard; Ostermann, Kai. *Temporal and spatial properties of a yeast multi-cellular amplification system based on signal molecule diffusion*. Sensors 13:14511-22, **2013**.

Jahn, Michael; Seifert, Jana; Hübschmann, Thomas; von Bergen, Martin; Harms, Hauke; Müller, Susann. *Comparison of preservation methods for bacterial cells in cytomics and proteomics*. Journal Of Integrated Omics 3:1-9, **2013**.

Jahn, Michael; Seifert, Jana; von Bergen, Martin; Schmid, Andreas; Bühler, Bruno; Müller, Susann. *Subpopulation-proteomics in prokaryotic populations*. Current Opinion in Biotechnology 24:79-87, **2013**.

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

Jahn, Michael; Vorpahl, Carsten; Türkowsky, Dominique; Lindmeyer, Martin; Bühler, Bruno;Harms, Hauke; Müller, Susann. *Accurate determination of plasmid copy number of flow-sorted cells using droplet digital PCR*. Analytical Chemistry 86:5969-76, **2014**.

Conference contributions

Oct 15-17, 2014	**DGFZ** – 24th Annual Conference of the German Society for Cytometry, Dresden, **presentation** 'Plasmid copy number variation in *Pseudomonas putida* analyzed by cell sorting and Digital Droplet PCR'
Jul 13-16, 2014	**ECB16** – 16th European Congress on Biotechnology, Edinburgh, **presentation** 'Population heterogeneity in *P. putida* analyzed on the single cell level using proteomics and digital PCR'
May 26-28, 2014	**DECHEMA** – Biomaterials Made in Bioreactors, Dresden, **poster** 'Plasmid copy number variation as a source for population heterogeneity'
Sep 23, 2013	**HiGRADE conference**, Leipzig, **presentation** 'To glow or not to glow: The manifold decisions of protein producing bacteria'
Mar 06-08, 2013	**RPP7** – 7th European Conference on Recombinant Protein Production, Ulm, **presentation** 'Deciphering heterogeneity: Proteomics on the sub-population level'
Jul 21-25, 2013	**FEMS** – 5th Congress of European Microbiologists, Leipzig, **poster** "Subpopulation-proteomics in prokaryotic populations"

Oct 10-12, 2012   **DGFZ** – 22nd Annual Conference of the German Society for Cytometry, Bonn, **presentation** 'Subpopulation-proteomics of bacterial cultures in biotechnology'

Jun 23-27, 2012   **CYTO 2012** – 27th Congress of the International Society for Advancement of Cytometry, Leipzig, **presentation** 'Storage methods for flow cytometric analysis of bacterial cells'

5.4 Glossary

1st generation biomass Agricultural biomass usable for human diet or animal feeding, such as starch or saccharose.

2nd generation biomass Agricultural biomass not usable for human diet or animal feeding, such as straw or wood, as opposed to →1st generation biomass.

(Bacterio-) Phage Virus with bacteria as host organisms. A self-replicative, packaged portion of DNA or RNA.

Biotransformation The conversion of one chemical compound into another by a biological catalysts such as an enzyme or ribozyme.

bp One base pair, the length unit of DNA.

Chaperone A protein helping other proteins to mature and fold correctly.

***cis-trans* isomerism** In organic chemistry, the position of side chains at a C double bond. Either the two side chains with highest priority are on the same (*cis*) or the opposite site (*trans*).

Digital PCR Next generation of quantitative PCR, where template molecules are discretely distributed across many small partitions and detected individually.

Discreteness In statistics, a distribution where all components must be integers (whole numbers).

Epoxidation Chemical reaction resulting in the formation of a heterocyclic 3-ring with two carbon and one oxygen atom.

Extrinsic noise Fluctuation in gene expression due to extracellular, 'environmental' perturbations.

Green Fluorescent Protein, GFP A protein naturally occurring in the jellyfish *Aequoria victoria*. Its unique reaction center is able to absorb light and emit fluorescence of a higher wavelength.

Growth rate In microbiology, the fraction of growth of a population per time unit.

Heterologous protein A protein not naturally present in the host organism where it is expressed.

Homologous recombination The site-specific excision or insertion of DNA fragments *in vivo*, catalyzed by specific enzymes (recombinases).

Intrinsic noise Fluctuation in gene expression due to the →stochastic nature of transcription and translation.

Operon In bacteria, an array of genes uniformly regulated by the same regulatory site (operator).

Open reading frame, ORF In genetics, the base pair sequence between start and stop codon of a gene. A shift of one or two bp will destroy the reading frame by changing the encoded amino acids.

OriV, oriT Origin of replication (vegetative) and origin of transfer, for plasmid maintenance and transduction, respectively.

Plasmid A circular, self-replicating portion of DNA.

Poisson distribution In statistics, a →discrete probability distribution to estimate a given number of events occurring in a fixed interval of time or space.

qRT-PCR Quantitative real time polymerase chain reaction. A technique, where the *in vitro* synthesis of DNA is monitored by fluorescent dyes, allowing quantification of the DNA template.

Reactive oxygen species Either highly reactive oxygen radicals with a free electron ($O_2 \cdot^-, OH \cdot$), or reactive oxygen donors (hydrogen peroxide $HOOH$).

Restriction, ligation The cutting and re-joining of DNA in molecular biology.

Saturated/unsaturated fatty acids Unsaturated fatty acids do contain C double bonds, saturated do not.

SEVA Standard European Vector Architecture, a plasmid repository.

Stochasticity The amount of randomness of a process, where single events cannot be predicted but only estimated by probability. Examples are Brownian motion or radioactive decay of isotopes.

Suicide vector A →plasmid carrying an origin of replication not functional in the target strain. This strain is only resistant to an antibiotic if the suicide vector and its resistance gene integrates into the genome.

Synthetic biology The rational design and implementation of novel functions or modules in biological systems.

Systems biology In molecular biology, the view of an organism as a whole as opposed to single genes or proteins.

Toxin-antitoxin system →Plasmid addiction system, where a stable toxin is degraded by an unstable antitoxin. Plasmid loss results in retention of the toxin and killing of the host cell.

Transcription factor A protein promoting or inhibiting transcription by binding an →operator region on the DNA.

Vector A →plasmid modified to serve as a genetic vehicle for storage and transformation of genes into host cells.

5.5 Acknowledgements

5.6 Supplementary material

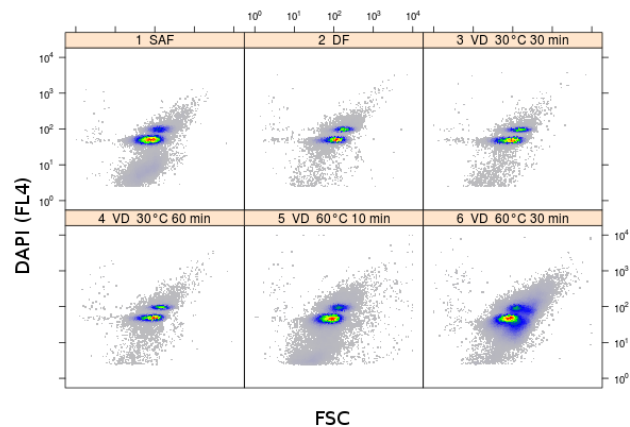## 5.6.1 Supplementary material for Publication 1

Figure 5.1: Supplementary Figure S1. Evaluation of vacuum drying (VD) conditions for *P. putida* KT2440 using flow cytometry. Depicted is forward scatter (FSC) versus DAPI fluorescence (FL4). Different drying temperatures (30°C, 60°C) and drying durations (10, 30 and 60 min) were tested as described in the materials and methods section, and the lower temperature combined with a moderate drying duration (30°C, 30 min) yielded the most distinct distribution. Sodium azide fixed (SAF) and deep frozen (DF) cells were used for comparison.

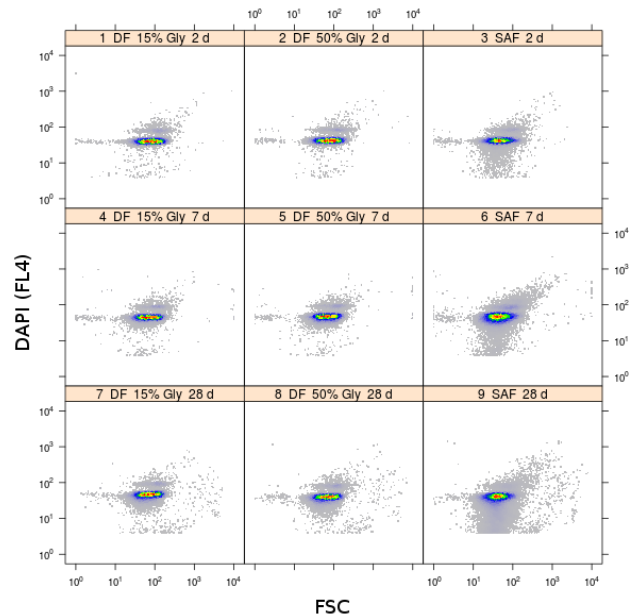**Figure 5.2:** Supplementary Figure S2. Evaluation of deep freezing (DF) conditions for *P. putida* KT2440 cells using flow cytometry. Depicted is forward scatter (FSC) versus DAPI fluorescence (FL4) for three different storage methods (DF with 15 and 50 % (v/v) glycerol in PBS as a cryoprotective agent, sodium azide fixation) and three storage durations (2, 7, 28 d). No significant difference was observed between 15 and 50 % glycerol.
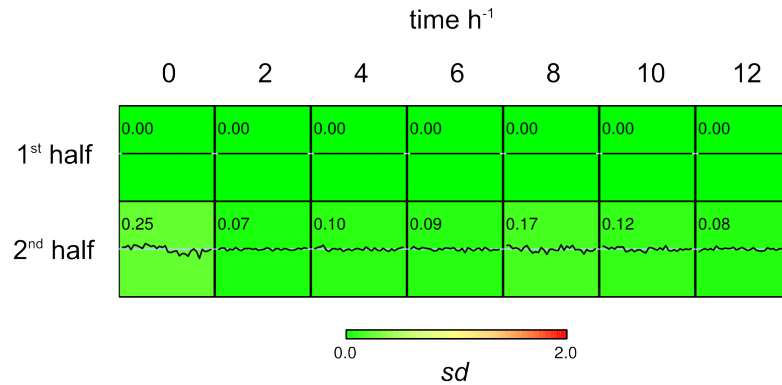
Figure 5.3: Supplementary Figure S3. Internal variation of identical samples when using FlowFP finger-printing. The raw data of the fresh samples for each time point (0, 2, 4, 6, 8, 10, 12 h) were *in silico* split in two halves and the similarity of both halves was computed. Depicted is the comparison of the first half versus itself and versus the second half. The standard deviation *sd* ranged from 0.07 to 0.25, which represents the internal variation of sample acquisition.
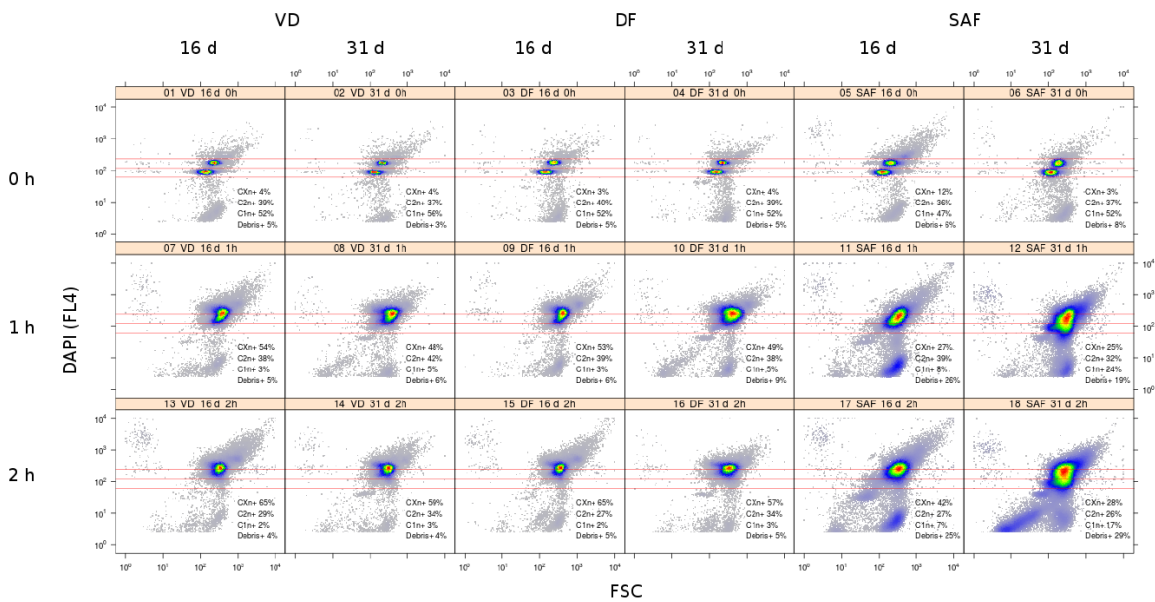
Figure 5.4: Supplementary Figure S4. Test of different storage methods for *E. coli* DH5α using flow cytometry. Vacuum drying (VD), deep freezing (DF) and sodium azide fixation (SAF) were applied to cells after 0, 1 and 2 h of incubation into growth medium. The prepared samples were stored for 16 or 31 days and then analyzed by flow cytometry as described in the materials and methods section. An increase of events below the C1n population is clearly visible for SAF treated samples (1, 2 h), particularly after 31 d of storage.

Figure 5.5: Supplementary Figure S5. Variability of protein number and quantity across replicates as identified by mass spectrometry. (A) The number of proteins detected by MS per replicate ranged from 611 to 832 (total: 971). Some replicates of 28 d stored VD and SAF treated samples showed a reduced protein number. The relative protein quantity of all proteins (without 1 % lower and upper quantile) was taken into account to calculate the similarity of a sample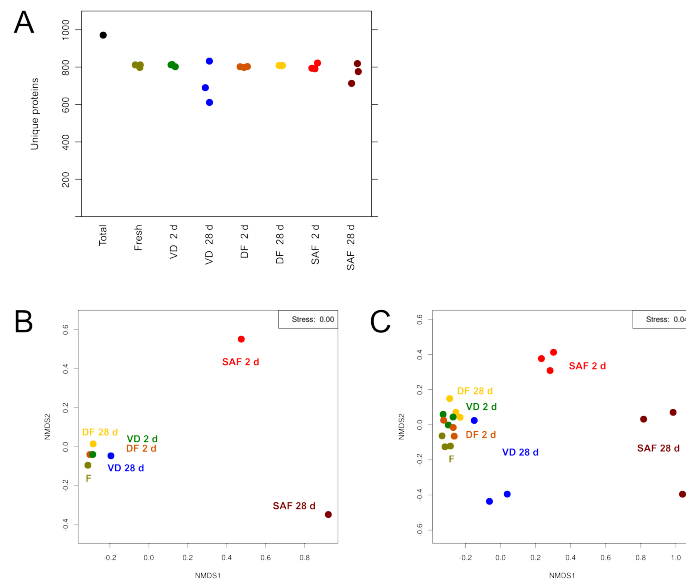 (B) or single replicate (C) towards each other. The distance, expressed as standard deviation, was visualized by non-metric multidimensional scaling (NMDS). The similarity to the fresh sample (F) was high for deep frozen (DF) and vacuum dried (VD) samples with the exception of lower similarity for 2 VD replicates after 28 d of storage. Sodium azide fixed samples (SAF) displayed the highest dissimilarity.
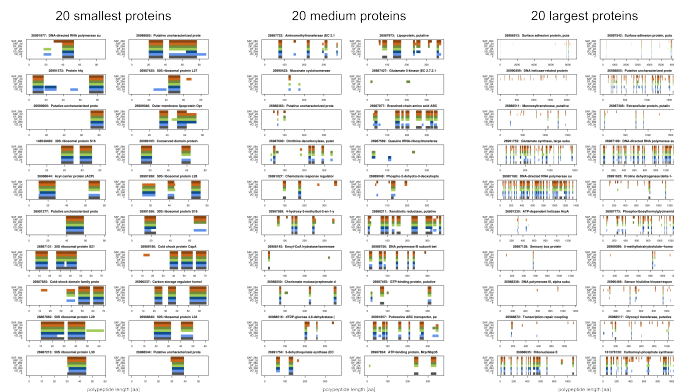
Figure 5.6: Supplementary Figure S6. Peptide coverage of selected proteins. For 60 proteins of different length (the 20 largest, 20 smallest and 20 of medium polypeptide length) all peptides, which were at least detected once by MS, were mapped onto the complete amino acid sequence. A different pattern in peptide coverage may indicate fragmentation or degradation at the termini of the polypeptide. Here, peptide coverage was highly similar across samples. Sample description: F – fresh, VD – vacuum drying, DF – deep freezing, SAF – sodium azide fixation.

Supplementary information

Supplementary information S1. MaxQuant settings for proteomics data analysis

The retrieved raw data were analyzed by MaxQuant version 1.2.2.5 [1] with the genome sequence of *P. putida* KT2440 as database. The settings for MaxQuant were the following: Peptide modifications given were methionine oxidation as variable and cysteine carbamidomethylation as fixed. Further settings were first search ppm of 20, main search ppm of 6, maximum number of modifications per peptide 5, max. missed cleavages 2 and a maximum charge for the peptide of 5. Parameters for the identification were a minimum peptide length of 5 amino acids, a false discovery rate for peptides, proteins and level of modification sites of 1%. A minimum of 2 unique peptides was required for protein identification. Apart from unmodified peptides only peptides with oxidized methionine and carbamidomethylized cysteine were used for quantification. Only unique or razor peptides were chosen for use in quantification. Further miscellaneous settings included re-quantify, match between runs (time window of 2 min), label free quantification and second peptides. To analyze the sample data label free quantification (LFQ) values were used.

Supplementary information S2. Sunburst treemaps

Treemaps are useful for visualization of hierarchically ordered data, such as functional annotation provided by KEGG. Both rectangular [2] and voronoi-based treemaps [3] have been used for data visualization, but here we report the first use of circular ("sunburst") treemaps for proteome data representation. Sunburst treemaps were created using a custom recursive function in R, where the size of each group sector corresponds to the number of proteins within the group, and the color encodes the output of an arbitrary function applied to the group subset, e. g. mean, median, variance or standard deviation. This *sectorplot()* function does not depend on R packages other than the basic ones. The source code for the function can be simply executed in R (tested with R version >= 2.14)

```
## PLOT FUNCTION
sectorplot    <- function(data, levels, exp, FUN, adjust.height=FALSE,
height.inner=0.1, height.initial=0.5/length(levels), range=c(-1,1), col=NULL,
border=NULL, lwd=1, labels=NA) {

# FUNCTION FOR DRAWING SECTOR
draw.sector <- function(height=0.1, a=1, segment=c(0, 2*pi), nv=30, border=NULL,
col=NA, lty=1, lwd=1, draw.label=FALSE, categ, ...) {
    z <- seq(segment[1], segment[2], length=nv + 1)
    xx <- c( a * cos(z), rev((a+height)* cos(z)))
    yy <- c(a * sin(z), rev((a+height) * sin(z)))
```

1

```
    polygon(xx, yy, border=border, col=col, lty=lty,
    lwd=lwd, ...)
    # draw labels
    if (draw.label) {
        if (a*cos(median(z))>=0) side=1 else side= -1
        lines(    x=c((a+height)*cos(median(z)), 0.8*cos(median(z)), 0.8*side),
            y=c((a+height)*sin(median(z)), 0.8*sin(median(z)), 0.8*sin(median(z))),
            col="grey")
        text(x=0.8*side, y=0.8*sin(median(z)), labels=substring(categ, 1, 15),
col="grey", pos=3+side, offset=0.2, cex=0.7)
    }
}

# CALCULATE SECTOR RANGE AND DRAW SECTOR
sector <- function(level, categ, rows.select, height.inner, height.actual,
expression) {
    hits <- which(data[[levels[level]]] %in% categ)
    # Color definition
    if (is.null(col)) {color <- colorRampPalette(c("#539FD4","#8BBF09","#F1DB5B"))
(length(levels))[[level]]}
    else {color <- colorRampPalette(col)(100)[1+((expression-range[1])/(range[2]-
range[1])*99)]}
    draw.sector(a=height.inner,
        height=height.actual,
        segment=c(hits[1], hits[length(hits)]+1)/nrow(data)*2*pi,
        border=border, lwd=lwd, categ=categ,
        col=color, draw.label=(!is.na(labels) & levels[level]==labels))
}

# CORE FUNCTION
recur <- function(level=1, rows.select=1:nrow(data), height.inner) {
    for (categ in unique(data[rows.select,levels[level]])) {
        # DRAW SINGLE SECTOR
        expression <- FUN(data[data[[levels[level]]]==categ, exp])
        if (adjust.height) {height.actual <- height.initial+
(height.initial*(expression/(range[2]-range[1])) )}
        else {height.actual <- height.initial}
        sector(level, categ, rows.select, height.inner, height.actual, expression)
        # CALL FUNCTION RECURSIVELY
        if (level!=length(levels)) { recur(level=level+1,
rows.select=which(data[[levels[level]]] %in% categ),
height.inner=height.inner+height.actual) }
    }
}

# FUNCTION CALL
par(mar=c(0,0,0,0))
plot(0, type="n", xlim=c(-1,1), ylim=c(-1,1), axes=FALSE)
recur(level=1, height.inner=height.inner)

# DRAW LEGEND
xcoord <- grconvertX(0, from="ndc", to="user")
ycoord <- grconvertY(1, from="ndc", to="user")
if (is.null(col)) {legend(xcoord, ycoord, legend=levels, pch=15, pt.cex=3,
col=colorRampPalette(c("#539FD4","#8BBF09","#F1DB5B"))(length(levels)), xpd=NA,
bty="n")}
else {legend(xcoord, ycoord, legend=seq(range[1],range[2], length.out=length(col)),
```

2

```
pch=15, col=col, pt.cex=3, xpd=NA, bty="n")}
}
```

The function takes the following arguments for (graphical) adjustments:

```
sectorplot(data, FUN, levels, exp, adjust.height=TRUE, height.inner=0.1,
height.initial=0.5/length(levels), range=c(-1,1), col=NULL, border=NULL,
lwd=1, labels=NA)
```

Argument	Usage
data	a data.frame object with hierarchical ordered data in columns
FUN	function applied to group sectors for color coding (like mean, median,...)
levels	character vector indicating column names of the data.frame
exp	character indicating the column with values, which are used by FUN
adjust.height	logical. Should the height of sectors be adjusted to exp values?
height.inner	numeric. Adjusts the "hole" of the innermost sector level
height.initial	numeric. Adjusts basic height of sectors
range	the range to which colors are mapped
col	colors to construct a color gradient (passed to colorRampPalette)
border	line color of sector borders
lwd	line width of sector borders
labels	character indicating which column to choose for labels

The function can be used according to the following self-contained example. First, we construct a data.frame and then visualize it using *sectorplot()*.
The function supports only hierarchical structures where each 'mother' group has a distinct set of 'children'. Ambiguous allocation of child groups may lead to unintended results.

```
dat <- data.frame(A=rep(1:2, each=6), B=rep(1:4, each=3), C=sample(1:12, 12))

sectorplot(dat,
    FUN=mean, levels=c("A","B","C"),
    exp="C", labels="C",
    range=c(0,12),
    col=c("lightgrey","darkgrey","royalblue","orange","red"),
)
```

3

References

1. J. Cox, M. Mann, Nat Biotechnol 26 (2008) 1367-1372. http://view.ncbi.nlm.nih.gov/pubmed/19029910

2. E.H. Baehrecke, N. Dang, K. Babaria, B. Shneiderman, BMC Bioinformatics 5 (2004) 84. http://view.ncbi.nlm.nih.gov/pubmed/15222902

3. D. Becher, K. Hempel, S. Sievers, D. Zühlke, J. Pané-Farré, A. Otto, S. Fuchs, D. Albrecht, J. Bernhardt, S. Engelmann, U. Völker, J.M. van Dijl, M. Hecker, PLoS One 4 (2009) e8176. http://view.ncbi.nlm.nih.gov/pubmed/19997597

4

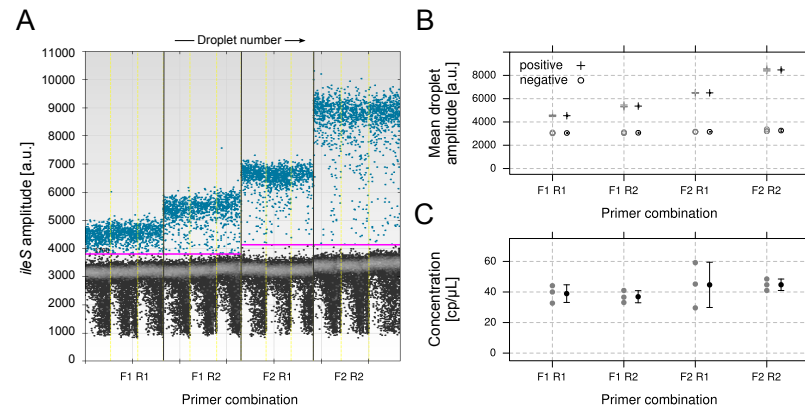## 5.6.2 Supplementary material for Publication 3

Figure 5.7: Supplementary Figure S1. Influence of primer combinations on *ileS* ddPCR. For each PCR reaction two different forward and reverse primers were tested in all four possible combinations. A.u., arbitrary units. (A) For the example of *ileS*, all combinations allowed differentiation of negative and positive droplets, but fluorescence amplitude was different. (B) Mean droplet amplitude for the four different combinations. (C) Despite this difference, the calculated copy number is identical for all tested combinations. Cp/µL, copies/µL. Grey symbols, replicate values. Black symbols, mean ± standard deviation.
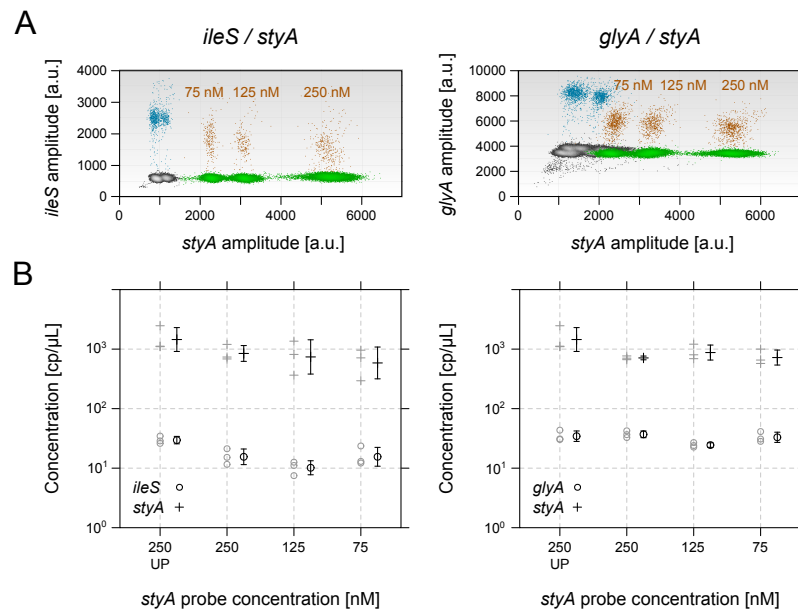
Figure 5.8: Supplementary Figure S2. Optimization of probe concentration in duplex reactions. Duplex ddPCR was performed with one genomic (*ileS* or *glyA*) and one plasmid marker (*styA*) using FAM- and HEX-labeled probes, respectively. (A) The styA probe concentration was optimized for best separation of negative (grey), single positive (blue, green) and double positive (brown) droplets. The scatter plot combines samples of all three probe concentrations and the respective clouds are indicated in the plot. A.u., arbitrary units. (B) The obtained copy numbers are similar for the different probe concentrations and also in comparison to the *styA* uniplex control (UP). Cp/µL, copies/µL. Grey symbols, replicate values. Black symbols, mean ± standard deviation.
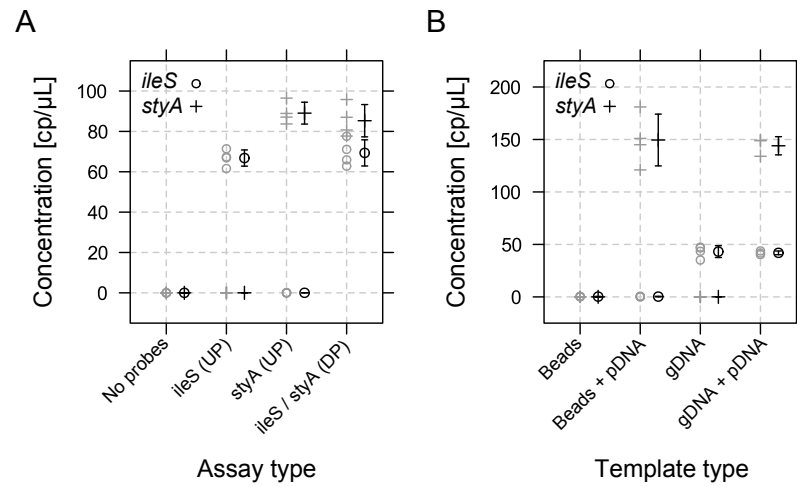
Figure 5.9: Supplementary Figure S3. Controls for duplex ddPCR. (A) Multiplexing controls for duplex ddPCR assay. Either nothing (no probes), *ileS* or *styA* in uniplex reactions (UP) or both markers together in a duplex assay (DP) were detected. Template is an artificial mixture of genomic and plasmid DNA. (B) Different mixtures of template were used to control reliability of the *ileS*/*styA* duplex assay. Beads, 1,000 sorted beads. gDNA, genomic DNA of *P. putida* KT2440. pDNA, plasmid pA-EGFP-B. Cp/μL, copies/μL. Grey symbols, replicate values. Black symbols, mean ± standard deviation.

Figure 5.10: Supplementary Figure S4. Scheme of cultivation with variable concentration of antibiotic. Transformants were grown on LB medium supplemented with 500 μg/mL kanamycin, and used to inoculate shaking flasks with minimal medium (MM). Cells were pre-cultivated twice for 16 h with variable amounts of kanamycin (0, 50, 500 μg/mL) and cultivated once without kanamycin but induction with 0.025 % DCPK.

Supporting Information for:

Accurate Determination of Plasmid Copy Number on the Single Cell Level using Droplet Digital PCR

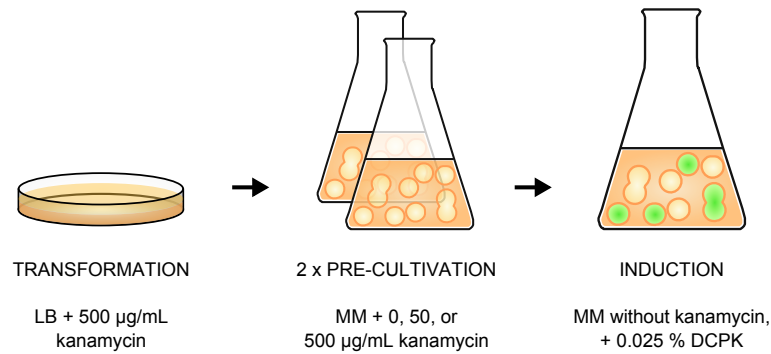*Michael Jahn[1], Carsten Vorpahl[1], Dominique Türkowsky[1], Martin Lindmeyer[2], Bruno Bühler[2], Hauke Harms[1], Susann Müller[1]*

[1]Helmholtz-Centre for Environmental Research – UFZ, Permoserstraße 15, 04318 Leipzig, Germany

[2]Laboratory of Chemical Biotechnology, Department of Biochemical and Chemical Engineering, TU Dortmund University, Emil-Figge-Str. 66, 44227 Dortmund, Germany

EXPERIMENTAL SECTION

Cultivation of bacteria

Pseudomonas putida KT2440 (obtained from DSMZ – German Collection of Microorganisms and Cell Cultures) was plated on LB medium[1] with 2 % agarose, incubated overnight, used for inoculation of 10 mL minimal medium in 100-mL shake flasks and cultivated at 30 °C on a rotary shaker (200 rpm). Baffled shake flasks with 25 mL medium were inoculated with a volume of overnight culture corresponding to an optical density at 600 nm ($OD_{600\,nm}$, Ø=0.5 mm) of 0.1. Cultivation media were LB without supplements or minimal medium (($NH_4)_2SO_4$ 2.2 g/L, $MgSO_4$ x $7H_2O$ 0.4 g/L, $CaCl_2$ x $2H_2O$ 40 mg/L, NaCl 20 mg/mL, Na_3-citrate x $2H_2O$ 15 mg/L, $FeSO_4$ x $7H_2O$ 10 mg/L, $ZnSO_4$ x $7H_2O$ 2 mg/L, $MnCl_2$ x $4H_2O$ 1 mg/L, $CuSO_4$ x $5H_2O$ 1 mg/L, $NiCl_2$ x $6H_2O$ 20 µg/L, $NaMoO_4$ x $2H_2O$ 30 µg/L, H_3BO_3 0.3 mg/L) supplemented with 5 g/L citrate as the carbon and energy source. If required, kanamycin (Cat-No. T832.3, Roth) was added as a 50 mg/mL stock solution in deionized H_2O (dH_2O) to a final concentration of 50-500 µg/mL and gene expression was induced with addition of dicyclopropylketone (DCPK, Sigma) to a final concentration of 0.025 %. Measurement of $OD_{600\,nm}$ and EGFP fluorescence (485 nm excitation/520 nm emission wavelength) during cultivation was performed in a Tecan GENios Plus spectrophotometer by sampling of 200 µL culture in a transparent 96-well plate. Samples for flow cytometry were taken by centrifugation of 1 mL cell suspension for 2 min at 8,000 x g and 4 °C. The supernatant was discarded and the cells re-suspended in 500 µL ice cold cryopreservation buffer as described in Jahn et al.[2]. Cell samples were stored at -20 °C until analysis.

1

DNA manipulation

Agarose gel electrophoresis and ligation was performed as described by Sambrook and Russel[1]. Enzymes (Phusion High-Fidelity Polymerase, T4 DNA ligase, restriction enzymes) and buffers were purchased from Thermo Scientific and oligonucleotides from Sigma Aldrich. Plasmids and DNA-fragments were isolated using the peqGOLD plasmid Miniprep Kit I (peqLab) and purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) according to the supplier's recommendation.

For construction of the plasmid pA-EGFP_B, the fusion of *styA* (oxygenase component of the styrene monooxygenase from *Pseudomonas* sp. strain VLB120[3]) and *EGFP* was performed via overlap extension PCR[4] using the primers F1-F4 and B1-B4 listed in Table S-1. The single genes *styA* and *styB* were amplified from genomic DNA using the primers F1/B1 and F3/B3 as the template and *EGFP* was amplified using the primers F2/B2 from plasmid pBSK-GFP1 (unpublished vector). To avoid steric hindrance, a linker sequence (GGC GGC GGC GGC GGC GGC GCC) was integrated between the C-terminus of StyA and the N-terminus of EGFP. After amplification and purification, the created fragment was cloned via blunt end ligation into pSMART-HCKan (Lucigen) and transformed into *E. coli* DH5α. Correctly ligated plasmids were isolated and used as PCR template. The resulting PCR product generated with primers F5 and B5 and plasmid pCom10[5] were digested with NdeI, purified, and ligated according to the supplier's recommendations. The correct insertion was verified by digestion with different restriction enzymes and DNA sequencing.

Electroporation

For electro-competent cells, a 100-mL volume of LB medium was inoculated with a *P. putida* KT2440 overnight culture and grown until the $OD_{600\,nm}$ reached 0.5. All cells were harvested by centrifugation (5 min, 5,000 x g, 4 °C), washed once with 5 mL ice cold dH_2O and once with 5 mL ice cold 15 % glycerol, finally taken up in 2 mL 15 % glycerol and stored in 40-µL aliquots at -80 °C. For electroporation, an aliquot of competent cells was thawed on ice, 100 ng plasmid DNA (pA-EGFP_B) was added and the suspension transferred to an electroporation cuvette (1 mm gap). Electroporation was done in a Bio-Rad MicroPulser with program Eco1 at 1.8 kV and the cells were transferred to 1 mL LB medium and incubated at 30 °C for 30 min. Transformants were obtained by selection on LB agar (2 %) plates supplemented with kanamycin (50 µg/mL). For *E. coli*, electroporation was performed as described by Sambrook and Russel[1] using an Equibio EasyjecT Prima electroporator with 2.5 kV.

Statistics

If not indicated otherwise, experiments were performed with two independent biological replicates, and all PCR experiments were performed with at least three technical replicates per condition at the stage of cell sorting. Repeatability and inter-assay variation were monitored using the following controls; sorted beads (no-template-control), sorted beads spiked with plasmid (no-genomic-DNA-control), and sorted wild type cells (no-plasmid-control). Data acquisition for ddPCR was performed with QuantaSoft v1.3.2.0 software (calibrated for FAM/HEX, manual gating), data were exported as text file and further analyzed using R v3.0.1. For each reaction volume of 20 µL up to 15,000 droplets were analyzed with an average droplet volume of ~ 0.9 nL[6]. The PCN was calculated as the ratio of plasmid marker copy number and genomic reference copy number per replicate (c_{pDNA}/c_{gDNA}), indicated is mean and standard deviation of all replicates per condition. Statistical power analysis was performed with the R package 'pwr' to determine the minimum number of replicates for significant detection of PCN differences (power=0.8). Outliers were not removed except for known pipetting errors.

2

Table S-1. *Oligonucleotide primers for construction of plasmid pA-EGFP_B.*

Primer	Characteristics	Primer / probe sequence
F1	StyA_Fusion_Fw	GAATAAGCAGATTGCAGGCGAGCTGGGTATTG
B1	StyA_Fusion_Bw	CTCCCTTACTAACCATGGCGCCGCCGCCGCCGCCGGCCGCGATAGTGGGTGCG
F2	EGFP_Fusion_Fw	CCCACTATCGCGGCCGGCGGCGGCGGCGGCGCCATGGTTAGTAAGGGAGAGG
B2	EGFP_Fusion_Bw	GCCTTGACCAGCGGAGCAATAGCGTTACTTGTACAGCTCGTCCATGCC
F3	StyB_Fusion_Fw	GCATGGACGAGCTGTACAAGTAACGCTATTGCTCCGCTGGTCAAGGC
B3	StyB_Fusion_Bw	TCTTCCGGTGATCGGCACAGAAAGGC
F4	StyA*_Amp	GGGTATTGCCGAGGTAACGGTAAAG
B4	StyB*_Amp	CCGGTGATCGGCACAGAAAGGCCT
F5	StyAB_pCom10_*Nde*I_	CGGC<u>CATATG</u>AAAAAGCGTATCGG
B5	StyAB_pCom10_*Nde*I_	CGGC<u>CATATG</u>TCAATTCAGTGGCAACG

Table S-2. Oligonucleotide primers and probes used for ddPCR.

Target	Marker	Gene ID	Amplicon Length	Primer / probe sequence	T_m / °C	Length / bp
Genome *P. putida*	*ileS*	PP_0603	120	5'-GGACAACCCATACAAGACC-3'	60.16	19
				5'-TCAAAGCACCAGTTCACC-3'	60.26	18
				5'-FAM-TCCGCGCCCTGGCCGA-BHQ-1-3'	72.69	16
	glyA	PP_0671	145	5'-CCAGAGTTCAAGGCTTAC-3'	57.38	18
				5'-CGGAGATTTCCTGCTTG-3'	57.68	17
				5'-FAM-ACCACCTGTTCCTGCTGTCGCT-BHQ-1-3'	70.96	22
Plasmid pA-EGFP_B	*styA*[7]	Genbank: AF03116	115	5'-GGCTGGTAGAGACGGTAG-3'	60.97	18
				5'-CTGAGGAGTTTGGTTATTTCG-3'	59.12	21
				5'-HEX-TGGGAGCCTTGAGATCACCGTAG-BHQ-1-3'	68.02	23
	oriT[5]	Genbank: AJ30208	104	5'-CAGGTGCGAATAAGGGAC-3'	60.25	18
				5'-GTAGACTTTCCTTGGTGTATCC-3'	60.33	22
				5'-HEX-CCTATCCTGCCCGGCTGACG-BHQ-1-3'	69.64	20

3

Table S-3. dMIQE check list for digital PCR experiments.

ITEM TO CHECK	IMPOR-TANCE	CHECK-LIST	COMMENT
EXPERIMENTAL DESIGN			
Definition of experimental and control groups	E	+	Induced / non-induced cells, fluorescent / non-fluorescent cells, beads
Number within each group	E	+	4 replicates per group
Assay carried out by core lab or investigator's lab?	D	+	Investigator's lab
Power analysis	D	+	R package 'pwr', power 0.8
SAMPLE			
Volume/mass of sample processed	E	+	1,000 sorted cells or beads
Microdissection or macrodissection	E	NA	
If frozen - how and how quickly?	E	+	-20 °C, after centrifugation
If fixed - with what, how quickly?	E	NA	
Sample storage conditions and duration (especially for formalin fixed/paraffin embedded samples)	E	+	-20 °C, 15 % glycerol in PBS, 1month
NUCLEIC ACID EXTRACTION			
Quantification—instrument/method	E	+	NanoDrop, spectrometrically
Storage conditions: temperature, concentration, duration, buffer	E	+	-20 °C, 1,000 cells/well in 7 µL dH2O + 1 µL sheath buffer
DNA or RNA quantification	E	NA	NanoDrop, only for purified DNA controls
Quality/integrity, instrument/method, e.g. RNA integrity/R quality index and trace or 3':5'	E	NA	
Template structural information	E	+	whole cell DNA
Template modification (digestion, sonication, preamplification, etc.)	E	NA	
Template treatment (initial heating or chemical denaturation)	E	+	Heating 95 °C, 5-20 min depending on template
Inhibition dilution or spike	E	+	Sorted different cell numbers as template (10-10,000), inhibition for ≥ 5,000 cells
DNA contamination assessment of RNA sample	E	NA	
Details of DNase treatment where performed	E	NA	
Manufacturer of reagents used and catalogue number	D	NA	
dPCR TARGET INFORMATION			
Sequence accession number	E	+	see Table S2
Amplicon location	D	+	see Table S2
Amplicon length	E	+	see Table S2
In silico specificity screen (BLAST, etc.)	E	+	
Pseudogenes, retropseudogenes or other homologs?	D	+	no homologs
Sequence alignment	D	+	PrimerBLAST
Secondary structure analysis of amplicon and GC content	D	-	only of probes and primers
Location of each primer by exon or intron (if applicable)	E	NA	
Where appropriate, which splice variants are targeted?	E	NA	
dPCR OLIGONUCLEOTIDES			
Primer sequences	E	+	see Table S2
RTPrimerDB Identification Number	D	NA	
Probe sequences	D	+	see Table S2
Location and identity of any modifications	E	+	3' FAM/HEX fluorophore, 5' BHQ1 quencher
Manufacturer of oligonucleotides	D	+	MWG Operon

4

Purification method	D	+	HPLC (MWG Operon)
dPCR PROTOCOL			
Reaction volume and amount of cDNA/DNA	E	+	20 µL, containing 8 µL with 1,000 cells as template
Primer, (probe), Mg++ and dNTP concentrations	E	+	Primer 900 nM, Probes 125 or 250 nM, Mg^{2+} and dNTP proprietary (Bio-Rad)
Polymerase identity and concentration	E	-	proprietary (Bio-Rad)
Buffer/kit catalogue number and manufacturer	E	+	2x ddPCR Supermix for Probes, catalog: 186-3010 (Bio-Rad)
Exact chemical constitution of the buffer	D	-	proprietary (Bio-Rad)
Additives (SYBR Green I, DMSO, etc.)	E	NA	
Plates/tubes manufacturer and catalog number	D	+	Cell sorting: 8 well strips, clear, flat-cap, catalog: G003-SF-I (Kisker Biotech). PCR: twin.tec 96 well plates, clear, semi-skirted, catalog: 0030 128 575 (Eppendorf)
Complete thermocycling parameters	E	+	2-step cycling protocol, see manuscript
Reaction setup, gravimetric or volumetric dilutions (manual/robotic)	D	+	Manual volumetric dilutions
Total PCR reaction volume prepared	D	+	20 µL
Partition number, individual partition volume	E	+	~ 20,000 droplets / sample, ~0.9 nL
Total volume of the partitions measured (effective reaction size)	E	+	~ 15,000 droplets * 0.9 nL = 13.5 µL
Partition volume variance/SD	D	-	Unknown
Comprehensive details and appropriate use of controls	E	+	Controls: NTC (beads), no-gDNA-control (spiked plasmid), no-plasmid-control (WT cells), multiplex/uniplex assays
Manufacturer of dPCR instrument	E	+	Bio-Rad QX100
dPCR VALIDATION			
Optimization data for the assay	D	+	Optimization: probe/primer sequence, concentration, annealing temp (qRT-PCR), 8-well-strips, multiplexing, cell number, 2 independent reference (*ileS*/*glyA*) and target genes (*styA*/*oriT*)
Specificity (when measuring rare mutations, pathogen sequences etc.)	E	NA	
Limit of detection of calibration control	D	+	LOD (3 x SD of blank) < 1 cp/µL
If multiplexing, comparison with singleplex assays	E	+	Yes, see Figure S-3
DATA ANALYSIS			
Mean copies per partition (Lambda Λ or equivalent)	E	+	$Λ = -\ln(1-k/n)$, k = positive partitions, n = all partitions; $Λ (500 < k < 14000) = -\ln(1-k/15{,}000) → 0.03 < Λ < 2.7$
dPCR analysis program (source, version)	E	+	Quantasoft v1.3.2.0 (Bio-Rad)
Outlier identification and disposition	E	NA	
Results of no-template controls	E	+	sorted beads, see Figure 4 or Figure S-3
Examples of positive(s) and negative experimental results as supplemental data	E	+	template controls, see Figure S-3
Where appropriate, justification of number and choice of reference genes	E	+	2 reference genes previously tested for qRT-PCR application
Where appropriate, description of normalization method	E	NA	No normalization used
Number and concordance of biological replicates	D	+	Usually 2 biological replicates, 4 technical replicates
Number and stage (RT or dPCR) of technical replicates	E	+	Technical replication during cell sorting
Repeatability (intraassay variation)	E	+	3 independent repetitions of induction experiment

5

Reproducibility (interassay/user/lab etc. variation)	D	-	Not performed
Experimental variance or CI	E	+	relative standard deviation (CV) = 5-15 %
Statistical methods used for analysis	E	+	Mean and standard deviation, power analysis, student's t-test
Data submission using RDML (Real-time PCR Data Markup Language)	D	-	not performed

E, Essential information; D, Desirable information; NA, not applicable; +/-, applied/not applied.

REFERENCES

(1) Sambrook, J.; Russell, D.W. *Molecular Cloning: A Laboratory Manual*; New York: Cold Spring Harbor Laboratory Press, 2001; Vol. *1*.

(2) Jahn, M.; Seifert, J.; Hübschmann, T.; von Bergen, M.; Harms, H.; Müller, S. *Journal Of Integrated Omics* **2013**, *2*, 25-33.

(3) Köhler, K.A.K.; Rückert, C.; Schatschneider, S.; Vorhölter, F.; Szczepanowski, R.; Blank, L.M.; Niehaus, K.; Goesmann, A.; Pühler, A.; Kalinowski, J.; Schmid, A. *J Biotechnol* **2013**, *168*, 729-730.

(4) Ho, S.N.; Hunt, H.D.; Horton, R.M.; Pullen, J.K.; Pease, L.R. *Gene* **1989**, *77*, 51-59.

(5) Smits, T.H.; Seeger, M.A.; Witholt, B.; van Beilen, J.B. *Plasmid* **2001**, *46*, 16-24.

(6) Dong, L.; Meng, Y.; Wang, J.; Liu, Y. *Anal Bioanal Chem* **2014**, *406*, 1701-1712.

(7) Panke, S.; Witholt, B.; Schmid, A.; Wubbolts, M.G. *Appl Environ Microbiol* **1998**, *64*, 2032-2043.

6

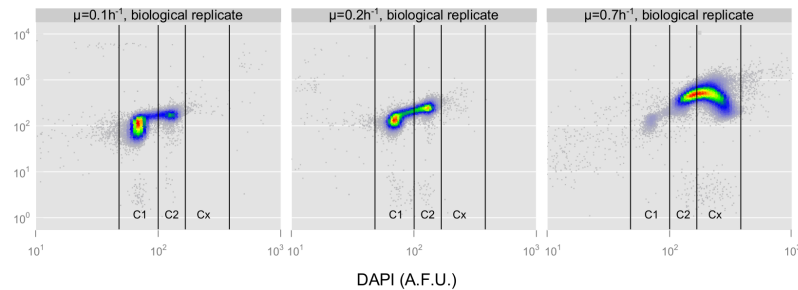## 5.6.3 Supplementary material for Publication 4

Figure 5.11: Supplementary Figure S1. 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, 0.1 h^{-1}, 0.2 h^{-1} and 0.7 h^{-1}. Cells of *P. putida* KT2440 grown at steady state conditions in chemostats were stained with DAPI and analyzed by flow cytometry. 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.
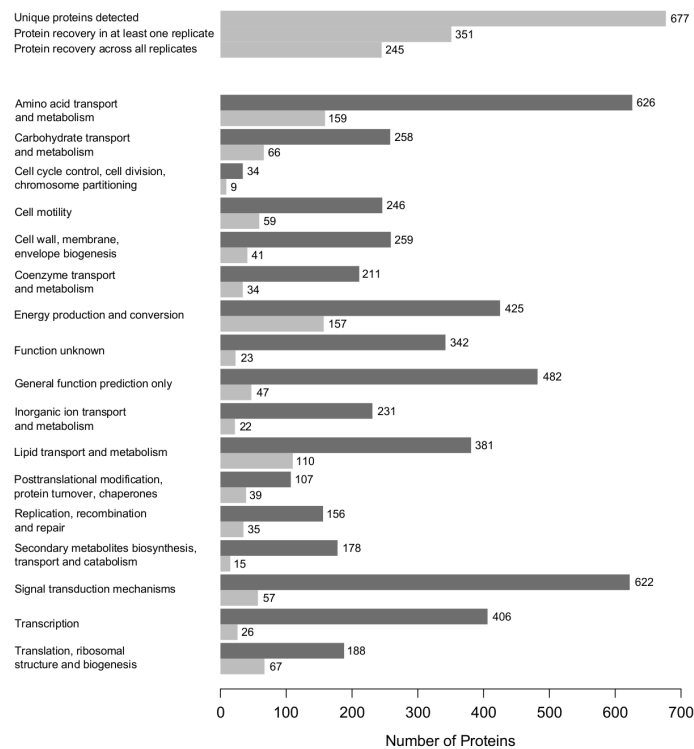

Figure 5.12: Supplementary Figure S2. Overview of the total protein detection and protein annotation. Overall, 677 unique proteins were identified, 351 proteins were detected in at least one replicate of all subpopulations and 245 proteins were found across all replicates. Functional annotation was carried 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).