DISSERTATION

Comprehensive analysis and prediction of large-scale transcript dynamics via the hybrid application of differential gene expression analysis and CFD

Von der Fakultät 4 - Energie-, Verfahrens- und Biotechnik der Universität Stuttgart zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von:
Julia WALTHER
aus Illertissen

Hauptberichter: Prof. Dr.-Ing. Ralf TAKORS
Mitberichter: Prof. Dr.-Ing. Michael SCHLÜTER

Tag der mündlichen Prüfung: 18.11.2021

Institut für Bioverfahrenstechnik

University of Stuttgart
Germany

2021

Für David

Declaration of Authorship

I, Julia WALTHER, declare that this thesis titled, "Comprehensive analysis and prediction of large-scale transcript dynamics via the hybrid application of differential gene expression analysis and CFD" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

Acknowledgements

I would like to use the following page to thank everyone without whom this work would hardly have been possible.

First and foremost I would like to express my gratitude to Prof. Dr.-Ing. Ralf Takors for giving me the opportunity to work on this interesting research topic. His valuable feedback, support and encouragement helped me to think from multiple perspectives and prepare this thesis. I am grateful for the trust and the freedom to develop the project according to my own ideas and to work on interesting side projects during my time at the IBVT. Furthermore, I want to thank Prof. Dr.-Ing. Michael Schlüter and Prof. Dr.-Ing. Joachim Groß for their interest in this research and their participation in the examination commitee as well as Prof. Dr.-Ing. Ulrich Nieken, Prof. Dr.-Ing. Eckart Laurien und Prof. Dr. Nicole Radde for their participation during the circulation procedure.

Secondly, I would like to thank my colleagues at the IBVT and especially my office colleagues Martin Ziegler and Maike Kuschel who offered a great professional and mental support during the ups and downs of my PhD. Many thanks go to Maike Kuschel and Flora Siebler who, together with me, uncovered and mastered the depths and mysteries of fluid flow simulation. Thank you both for the support, the good ideas and the time for the seemingly endless discussions. Special thanks go to my project partner Martin Ziegler. Through your support I was able to develop myself and the work considerably. Despite some tough phases, we managed to successfully complete the project and I am grateful I have had such a competent and professional colleague by my side during this time. I am grateful to Dr. Alexey Lapin, who has greatly enriched the work with his comprehensive and deep understanding of fluid mechanics and numerical simulation. I also want to thank the permanent staff at the IBVT for their support either in the lab or with administrative affairs: Ms. Reu with her warm and open manner, Andreas Freund and Alexander Dietrich with her helpful and competent support in the lab.

I also want to thank Vikas and Natascha with whom I was fortunate enough to work with. Thank you for the time and effort you put into your projects which made it easy and enjoyable to work with you.

I am grateful to my students Moritz Wild, Kimberly Fichtel, Janina Faiss and Matthias Konstantin for their enthusiasms and great ideas. I am glad to have been your colleague (even if only for a short amount of time) Moritz and wish you all the best for your upcoming journey at the IBVT.

Finally, my biggest thanks go to my family who supported me my whole life, which has enabled me to do many things. My parents and my brother who trusted in my abilities and my best friend and husband David who inspires me to take challenges and encourages me to develop my full potential. No matter what challenges I face, I always know I can count on you. I am looking forward to our future together and what is still waiting for us on this path. Thank you!

Contents

List of Figures

List of Tables

List of Abbreviations

AA	**A**mino **A**cid
aa	**A**mino **A**cyl
ADP	**A**denosine **D**i**P**hosphate
AI-2	**A**uto**I**nudcer-2
AMP	**A**denosine **M**ono**P**hosphate
ATP	**A**denosine **T**ri**P**hosphate
BH	**B**enjamini-**H**ochberg
bp	**B**ase**P**air
cDNA	**c**omplementary **DNA**
CDW	**C**ell **D**ry **W**eight
CFD	**C**omputational **F**luid **D**ynamics
COG	**C**lusters of **O**rthologous **G**roups
DEG	**D**ifferentially **E**xpressed **G**ene
DNA	**D**eoxyribo**N**ucleic **A**cid
DNS	**D**irect **N**umerical **S**imulation
DO	**D**issolved **O**xygen
DRW	**D**iscrete **R**andom **W**alk
EL	**E**uler-**L**agrange
FC	**F**old-**C**hange
FDA	**F**ood and **D**rug **A**dministration
FDR	**F**alse **D**iscovery **R**ate
GAM	**G**rowth **A**ssociated **M**aintenance
GEO	**G**ene **E**xpression **O**mnibus
GO	**G**ene **O**ntology
GTP	**G**uanosine **T**ri**P**hosphate
GRAS	**G**enerally **R**ecognized **A**s **S**ave
GTP	**G**uanosine **T**ri**P**hosphate
IB	**I**ndutsrial **B**iotechnology

IU	**I**nternational **U**nits
KOH	Potassium **H**ydr**O**xide
LES	**L**arge **E**ddy **S**imulation
LDA	**L**aser-**D**oppler **A**nemometry
MRF	**M**ulti **R**eference **F**rame
mRNA	**m**essenger **RNA**
MW	**M**olecular **W**eight
NAM	**N**on-growth **A**ssociated **M**aintenance
NGS	**N**ext **G**eneration **S**equencing
NMP	**N**ucleotide **M**ono**P**hosphate
NSE	**N**avier **S**tokes **E**quation
Nt	**N**ucleot**i**des
NTP	**N**ucleotide **T**ri**P**hosphate
ODE	**O**rdinary **D**ifferential **E**quation
PBE	**P**opulation **B**alance **E**quation
PBM	**P**opulation **B**alance **M**odel
PC	**P**rinicpal **C**omponent
PCA	**P**rinicpal **C**omponent **A**nalysis
PCD	**P**rogrammed **C**ell **D**eath
PCR	**P**olymerase **C**hain **R**eaction
PDO	1,3-textbfPropane**D**i**O**l
PFR	**P**lug **F**low **R**eactor
Q	Phred **Q**uality Score
RANS	**R**eynolds-**A**veraged **N**avier **S**tokes
rlog	**R**egularized-**L**ogarithm
RIB	**Rib**osome
RIN	**R**NA **I**ntegrity **N**umber
RLE	**R**elative **L**og **E**xpression
RNA	**R**ibo**N**ucleic **A**cid
RNAP	**RNA P**olymerase
RNG	**Re**n**o**rmalization **G**roup
rRNA	**r**ibosomal **RNA**
RSM	**R**eynolds **S**tresses as Second Order Closure **M**odel
S	**S**preading coefficient
SM	**S**liding **M**esh

SNP	**S**ingle **N**ucleotide **P**olymorphism
SDS	**S**hine-**D**algarno **S**equence
SQBC	**S**tochastic **Q**uality-**B**ased **C**lustering
sRNA	**s**mall **RNA**
SSA	**S**tochastic **S**imulation **A**lgorithm
STR	**S**tirred **T**ank **R**eactor
TA	**T**oxin/**A**ntitoxin
TC	**T**rans**c**ription
TL	**T**rans**l**ation
TMM	**T**rimmed **M**ean of **M**-values
TPM	**T**ranscripts **P**er **M**illion
tRNA	**t**ransfer **RNA**
VOF	**V**olume **o**f **F**luid

List of Symbols

Sign	Description	Unit
a	Distance	m
$a(V, V')$	Aggregation kernel	$m^3\,s^{-1}$
c	Speed of sound in medium	$m\,s^{-1}$
c_S, c_{O_2}	Substrate, Oxygen concentration	$mol\,L^{-1}$
$c^*_{O_2}$	Oxygen saturation concentration	$mol\,L^{-1}$
C	Cluster	
$C_{1\epsilon}, C_{2\epsilon}, C_\eta$	Empirical constants	
D	Dilution rate	h^{-1}
D_i	Diffusion coefficient	$m^2 s^{-1}$
d_i	Particle diameter	m
d_{32}	Sauter diameter	m
f	Fading genes	
f_i	External forcing term	
$g(V')$	Break up frequency	s^{-1}
$gene$	Gene	
g_i	Gravitational acceleration	$m\,s^{-2}$
G_k	Generation term of turbulent kinetic energy	
H	Working hypothesis	
H_0	Null hypothesis	
k	Turbulent kinetic energy	$m^2\,s^{-2}$
k_{deg}	Protein degradation rate	h^{-1}
k_La	mass transfer coefficient	h^{-1}
K_S	Substrate specific uptake constant	$g\,L^{-1}$
$L_{mRNA,i}$	Length of mRNA strand per gene i	nt
m_{ATP}	ATP consumption rate for maintenance demand	$mmol\,g^{-1}\,h^{-1}$
m_S	Pirt's maintenance coefficient	$g\,g^{-1}h^{-1}$
$M_{i,k}$	Interphase momentum exchange	

Sign	Description	Unit
Ma	Mach number	
max	Maximum number	
$Minsize$	Minimum number of observations per clusters	
MW	Molecular weight	$\mathrm{g\,mol^{-1}s}$
n	Number	
$n(V,t)$	Number density	$\mathrm{m^{-3}}$
N	Agitation speed	$\mathrm{s^{-1}}$
$Ntry$	Number of trials per iteration	
p	Pressure	$\mathrm{kg\,m^{-1}\,s^{-2}}$
$protein$	Bulk protein	
P	Power	$\mathrm{W\ (J\,s^{-1})}$
q_{ATP}	total cellular ATP consumption rate	$\mathrm{mmol\,g^{-1}\,h^{-1}}$
$q_{ATP,eGFP}$	ATP consumption rate for eGFP production	$\mathrm{mmol\ g^{-1}h^{-1}}$
$q_{\mathrm{s,max}}$	Maximum substrate uptake rate	$\mathrm{g\,(g\,h)^{-1}}$
Re	Reynolds number	
r	Distance of a gene from its cluster	
r_c	Capillary radius	m
r_{1-5}	Curvature radius	m
S	Number of starvation passages	
S_i	Silhouette coefficient of cluster i	
$S_{i,j}$	Strain rate	$\mathrm{s^{-1}}$
Sc_t	Turbulent Schmidt number	
t	Time (general)	s
T	Temperature	°C
T_i	Torque	$\mathrm{kgm^2s^{-2}}$
u	Fluid velocity	$\mathrm{m\,s^{-1}}$
u_i	Turbulent fluctuations	$\mathrm{m\,s^{-1}}$
V	Volume	$\mathrm{m^3}$
x	Location on DNA/ RNA strand	nt
x_1, x_2	Gene 1, gene 2	
x,y,z	Spatial coordinates	m
$Y_{ATP,S}$	ATP produced per fully oxidized substrate	$\mathrm{mmol\ mmol^{-1}}$
Y_{XS}	Biomass Yield	$\mathrm{g\,g^{-1}}$
$Y_{XS,true}$	True Biomass Yield	$\mathrm{g\,g^{-1}}$

Greek Symbols

α	Volume fraction	%
$\beta(V\|V')$	Breakage pdf	
δ_{ij}	Kronecker delta	
ϵ	Turbulent dissipation rate	$s^2\,m^{-3}$
η	Dynamic viscosity	$kg\,m^{-1}\,s^{-1}$
Γ_k	Source/Sink term	
μ	Growth rate	h^{-1}
μ_{lam}	Laminar dynamic viscosity	$kg\,m^{-1}\,s^{-1}$
∇	Nabla operator (the derivative in all spacial directions)	
ν	Kinematic viscosity	$m^2\,s^{-1}$
ρ	Density	$kg\,m^{-3}$
$\sigma_{i/j}$	Surface tension of phase i/j mixtures	$N\,m^{-1}$
$\sigma_k, \sigma_\epsilon$	Turbulent Prandtl number	
τ	Time scale	s

Indices

Sign	**Description**
crit	Critical
deg	Degraded
i	Iteration index
k, j	Data points
lg	Lagrange
max	Maximum
med	Median
min	Minimum
mix	Mixing
nucleotide	Nucleotide monophosphates
RIB	Active Ribosome
RNAP	Active RNAP
s	Substrate
t	Turbulent

Abstract

The growth in markets for recombinant proteins and a growing demand for biotechnological alternatives to chemical synthesis is placing increased pressure on efficient high-capacity production. Increased capacity is currently limited not only by insufficient scale-up techniques but also by non-robust production hosts that place limits on the process space-time yield. Heterogeneous reactor environments due to insufficient mixing and unsatisfactory mass transfer lead to frequent genomic perturbations, resulting in a decline in productivity. Recent studies have focused on computational fluid dynamics (CFD) to investigate large-scale mixing issues, which are often related to nutrient gradient formation. As numerical methods and computational power have improved over time and more complex and realistic models have been developed, experimental approaches have evolved as well. The latest generation of sequencing techniques allows for in-depth analysis of microbial behavior in scale-down setups under various conditions that support robust strain design. In the process, the attempt is made with scale-down reactors to recreate the observed conditions on a large scale.

While many researchers focus on one of the two approaches, this thesis provides methodologies to combine the microbial response derived from transcriptomic analysis with the substrate concentration distribution observed in a large-scale scenario. A framework is presented in which the results of transcriptomic analysis are used to create a better picture of gene expression dynamics under fluctuating substrate conditions derived from CFD simulations. The focus in this project is on *Escherichia coli* as the main production host but the methods presented can be just as well applied to other organisms and process parameters.

Additionally, to extend the single-phase representation of a large-scale environment to an industrially relevant multiphase scenario, preliminary investigations were performed for oil-in-water dispersions and the effect on the oxygen mass transfer coefficient ($k_L a$).

Although both research approaches typically only consider no more than two limitations simultaneously, the tools presented, for both the experimental and numerical approaches, allow for a rethinking of scale-up methods and the adaptation of techniques to

newly acquired knowledge.

Three case studies were performed, using the methods described above. At first a case study was conducted to determine whether a change in microbial behavior was due to engineered chassis design, rather than stress-induced mutations. For this purpose, a DNA-Seq analysis framework was established to identify mutations that could be harmful in relation to the desired behavior. A strain named *E. coli* RM214 was constructed, for use as a robust phenotype in heterogeneous fermentations. This performance of this strain was evaluated via fermentation in continuous cultivations in an STR-PFR scale-down reactor, with repeatedly induced glucose starvation stresses. *E. coli* RM214 showed beneficial behavior in the face of fluctuating glucose conditions by producing a 43 % higher eGFP yield after 28 h of cultivation. No modulation of cellular regulation in genes involved in the stringent response or the general stress response at any growth rate were found, which confirmed that the beneficial performance was due to engineered changes, rather than stress.

Transcriptomic analysis was used to obtain more detailed information on microbial behavioral changes and adaptation strategies in heterogeneous environments. In so doing, a transcriptomic analysis workflow was initially used to analyze adaptation strategies of a newly engineered *E. coli* strain with a modulated stringent response, named *E. coli* SR, which was optimized for large-scale conditions. An STR-PFR (stirred tank reactor - plug flow reactor) scale-down device was used to mirror fluctuating nitrogen conditions on a large scale. The strain showed no ppGpp-mediated stress response, while maintaining fully functional ammonium uptake and biomass formation. Additionally, *E. coli* SR showed a substantially reduced short-term transcriptional response, compared with *E. coli* MG1655, and adaptation was accomplished via negative regulation of transcription, translation and cell division. These results show that locally induced stress responses propagating through the bioreactor do result in reduced and coordinated response of genes in *E. coli* SR, which validates the intention behind the engineering of the chassis.

The transcriptomic data analysis workflow was subsequently extended via the use of glucose starvation data in the same experimental setup, to identify gene expression dynamics and adaptation strategies over 28 h of repeated glucose starvation. Ordinary differential equations (ODEs) in combination with stochastic elements were used to develop an agent-based model covering transcription and translation processes. In combination with a fluid-flow simulation of a 54 m^3 bioreactor, it was found that cellular ATP demand rises to 30 - 40 % of growth-decoupled maintenance (NGAM) demand, which may limit

ATP-intensive production formation accordingly. Furthermore, spatial analysis of individual transcriptional patterns reveals mRNA up-regulation with hot spots of 50 - 80 % in the upper region of the bioreactor. After 4.2 h cells adapt to environmental changes but still have to withstand an additional 6 % in NGAM demand.

The effect on bubble surface tension and size distribution, as well as oxygen mass transfer, was investigated in experiments via the extension from a single-phase to a multiphase scenario including safflower oil. In addition, the impact of the medium (water, complex glucose medium) and impeller setup on the aforementioned parameters was considered. While safflower oil reduces surface tension from 71 $\mathrm{mN\,m^{-1}}$ to around 61 $\mathrm{mN\,m^{-1}}$ in water, no additional effect was measured in the complex glucose medium. For all setups and parameters investigated, reduced surface tension resulted in smaller bubble sizes. However, $k_\mathrm{L}a$ measurements showed the opposite of the anticipated effect. Regardless of the medium or impeller setup used, $k_\mathrm{L}a$ values never exceeded the values of oil-free systems, and were diminished by up to 50 %.

In conclusion, the methods outlined above provide suitable tools for the incorporation of both physical and biological performance tests into scale-up processes. The proposed workflow offers a comprehensive approach, starting with the analysis of microbial response, taking into account nutrient gradient formation, and arriving at the effect prediction on a microorganism in terms of ATP in a large-scale process. Therefore, by delivering numerical performance-test tools, this research is contributing to a better understanding of the complex biological response toward insufficiently scaled up bioprocesses.

Zusammenfassung

Wachsende Märkte für rekombinante Proteine und eine steigende Nachfrage nach biotechnologischen Alternativen zur chemischen Synthese erhöhen den Druck auf eine effiziente Produktion mit hoher Kapazität. Die Steigerung der Kapazität wird dabei nicht nur durch unzureichende Scale - up Techniken begrenzt, sondern auch durch nicht robuste biologische Produzenten, welche die Raum - Zeit Ausbeute des Prozesses begrenzen. Heterogene Reaktorbedingungen aufgrund inhomogener Durchmischung und unzureichenden Stofftransfers führen zu häufigen genomischen Störungen bei Mikroorganismen und damit zu einer verschlechterten Produktivitätsleistung. Neueste Studien konzentrieren sich auf die numerische Strömungsmechanik (CFD), um großräumige Vermischungsprobleme zu untersuchen, die sich oft auf die Bildung von Nährstoffgradienten beziehen. Neben der Verbesserung der numerischen Methoden und der Rechenleistung in den letzten Jahren und der Entwicklung von komplexeren und realistischeren Modellen wurden auch die experimentellen Ansätze verbessert. Sequenzierungstechniken der nächsten Generation ermöglichen eine tiefgreifende Analyse des mikrobiellen Verhaltens in Scale - Down Reaktoren unter verschiedenen Bedingungen. Dabei versuchen Scale - Down Reaktoren, die beobachteten Bedingungen im großen Maßstab nachzubilden.

Während sich viele Forscher auf eines der beiden Themen konzentrieren, bietet diese Arbeit Methoden, um die aus der Transkriptomanalyse abgeleitete mikrobielle Reaktion mit einer Substratkonzentrationsverteilung eines Industriereaktors zu kombinieren. Es werden Techniken vorgestellt, basierend auf transkripto- mischen Analysen, um ein besseres Bild der Genexpressionsdynamik unter fluktuierenden Substratbedingungen zu erstellen, welche aus CFD - Simulationen abgeleitet wurden. Der Fokus in diesem Projekt liegt auf *Escherichia coli* als Hauptproduktionswirt, aber die vorgestellten Methoden können ebenso auf andere Organismen und Prozessparameter angewendet werden.

Um die einphasige Darstellung einer großskaligen Umgebung auf ein industrie- relevantes Mehrphasenszenario zu erweitern, wurden zusätzlich Voruntersuchung- en für Öl - in - Wasser Dispersionen und den Einfluss auf den Sauerstoff Massentransferkoeffizienten ($k_L a$) durchgeführt.

Obwohl in beiden Forschungsbereichen typischerweise nicht mehr als zwei Limitationen parallel betrachtet werden, erlauben die vorgestellten Werkzeuge, sowohl experimentell als auch numerisch, das Überdenken von Scale - up Methoden und die Anpassung der Techniken an neu gewonnene Erkenntnisse.

Unter Verwendung der etablierten Methoden wurden drei Fallstudien durchgeführt. Dabei wurde zuerst die Frage beantwortet, ob die mikrobielle Verhaltensänderung durch die gezielte Veränderung des Genoms und nicht durch stressinduzierte Mutationen erfolgt. Zu diesem Zweck wurde eine DNA-Sequenzierung verwendet, um potenziell schädliche Mutationen in Bezug auf das gewünschte Verhalten zu identifizieren. Ein Stamm, genannt *E. coli* RM214, wurde mit dem Ziel konstruiert, einen robusten Phänotyp unter heterogenen Fermentationsbedingungen zu erhalten. Die Leistung dieses Stammes wurde durch Fermentation in kontinuierlichen Kulturen in einem STR - PFR Scale - Down Reaktor bewertet, während wiederholter Glukose - Hunger Stress induziert wurde. *E. coli* RM214 zeigte ein vorteilhaftes Verhalten gegenüber schwankenden Glukosebedingungen, indem er nach 28 h Kultivierung eine 43 % höhere eGFP - Ausbeute produzierte. Es wurde keine Modulation der zellulären Regulation in Genen gefunden, die an der stringenten Reaktion oder der allgemeinen Stressreaktion bei jeder Wachstumsrate beteiligt sind, was die vorteilhafte Leistung in Bezug auf gezielte genomische Anpassungen und nicht auf Stress bestätigt.

Um detailliertere Informationen zu der mikrobiellen Verhaltensänderungen und Anpassungsstrategien in heterogenen Umgebungen zu erhalten wurde die transkriptomische Analyse verwendet. Hierbei wurde ein neu entwickelter *E. coli* Stamm mit veränderter stringenter Reaktion, genannt *E. coli* SR untersucht, welcher für großtechnische Bedingungen optimiert wurde. Ein STR - PFR (Stirred Tank Reactor - Plug Flow Reactor) Scale - Down Apparat wurde verwendet, um schwankende Stickstoffbedingungen im Großmaßstab wieder zu spiegeln. Der Stamm zeigte keine ppGpp-vermittelte Stressreaktion, während die Ammoniumaufnahme und Biomassebildung voll funktionsfähig blieben. Zusätzlich zeigte *E. coli* SR eine wesentlich reduzierte kurzfristige transkriptionelle Reaktion im Vergleich zu *E. coli* MG1655 und die Anpassung erfolgte über eine negative Regulation von Transkription, Translation und Zellteilung. Diese Ergebnisse zeigen, dass lokal induzierte Stressreaktionen, die sich durch den Bioreaktor ausbreiten, zu einer reduzierten und koordinierten Expression von Genen in *E. coli* SR führen, was die Intention der genomischen Modifikation validiert.

Die transkriptomische Datenanalyse wurde anschließend durch die Verwendung von

Glukose - Hunger Daten im gleichen Versuchsaufbau erweitert, um Genexpressionsdynamiken und Anpassungsstrategien über 28 h wiederholten Glukose - Hungers zu identifizieren. Gewöhnliche Differentialgleichungen (ODEs) in Kombination mit stochastischen Elementen wurden verwendet, um ein agentenbasiertes Modell zu entwickeln, das Transkriptions- und Translationsprozesse umfasst. Gekoppelt mit einer Strömungssimulation eines 54 m^3 Bioreaktors wurde festgestellt, dass der zelluläre ATP-Bedarf zwischen 30 - 40 % des wachstumsentkoppelten Erhaltungsbedarfs (NGAM) ansteigt. Darüber hinaus zeigt die räumliche Analyse einzelner Transkriptionsmuster eine mRNA-Hochregulation mit Hot Spots von 50 - 80 % im oberen Bereich des Bioreaktors. Nach 4.2 h passen sich die Zellen an die Umweltveränderungen an, müssen aber immer noch zusätzliche 6 % NGAM Nachfrage stemmen, was zur Limitation ATP intensiver Produktbildungen führen kann.

Durch die Erweiterung des einphasigen auf ein mehrphasiges Szenario mit Distelöl wurde die Auswirkung auf die Oberflächenspannung und Größenverteilung der Blasen, sowie auf den Sauerstoff Massentransfer experimentell untersucht. Zusätzlich wurde der Einfluss des Mediums (Wasser, komplexes Glukosemedium) und des Rühreraufbaus auf die genannten Parameter betrachtet. Während Distelöl in Wasser die Oberflächenspannung von 71 mN m^{-1} auf etwa 61 mN m^{-1} reduziert, wurde in komplexem Glukosemedium kein zusätzlicher Effekt gemessen. Für alle untersuchten Setups und Parameter führte eine reduzierte Oberflächenspannung zu kleineren Blasengrößen. Die $k_L a$ Messungen zeigten jedoch das Gegenteil des erwarteten Effekts. Unabhängig vom verwendeten Medium oder Rührer Setup überstiegen die $k_L a$ Werte nie die Werte von ölfreien Systemen und waren bis zu 50 % vermindert.

Zusammenfassend lässt sich sagen, dass die vorgestellten Methoden geeignete Werkzeuge für die Einbindung sowohl physikalischer als auch biologischer Leistungstests in Scale-up-Prozesse bieten. Der vorgeschlagene Arbeitsablauf bietet einen umfassenden Ansatz, der mit der Analyse der mikrobiellen Reaktion unter Berücksichtigung der Bildung von Nährstoffgradienten beginnt und mit der Wirkungsvorhersage auf einen Mikroorganismus in Bezug auf ATP in einem großtechnischen Prozess endet. Daher trägt diese Arbeit durch die Bereitstellung numerischer Leistungstest Methoden zu einem besseren Verständnis der komplexen biologischen Reaktionen in heterogenen Bioprozessen bei.

Chapter 1

Introduction

Modern industrial biotechnology (IB) emerged as a field in the 1970s and since then, has served as a sustainable alternative to the chemical industry. Detergent enzymes provide an example of an early commercial product, and still account for 30 % of an industrial enzyme market worth around $ 5.5 billion annually. Regarding the global market for industrial enzymes, a recent report published by BBC Research states that it should reach $ 7.0 billion by 2023, at a compound annual growth rate of 4.9 % for the period 2018-2023 (Research, 2018). A major contributor to the consistent growth of the IB market is the diversity of its products. Biotechnology products range from household cleaning products to food, pharmaceuticals and many other product types. Within this huge product spectrum, *Escherichia coli* is one of just a few main hosts used to produce recombinant proteins (Ferrer-Miralles et al., 2009; Yim et al., 2011). Many molecular toolboxes are available for *E. coli*, which facilitates the construction of expression systems suitable for high-yield enzyme production.

Industrial biotechnology is promoted as a promising key technology in numerous fields of business, research, and society with the aim to replace or at least complement chemical production (Fröhling et al., 2020). However, since this technology only recently emerged and due to the complex interplay between biology and process engineering, there is much room for further optimization. It is imperative to improve existing biotechnology processes and develop new methods in order to maintain the growth and competitiveness of this industry, in accordance with the goal of achieving a more sustainable future (Heinzle et al., 2007).

1.1 Challenges of scale-up

Low-cost, high-quality products are usually produced using large-scale processes for maximum efficiency. However, these processes are typically developed on a laboratory scale,

and the transfer to larger scales is often accompanied by a decrease in performance. Insufficient mixing and unsatisfactory mass transfer inside the reactor may lead to nutrient gradients triggering frequent genomic perturbations in production hosts, which result in reduced product yields. New scale-up strategies based on knowledge of local reactor conditions and their impact on microorganisms must be developed. Therefore, it is essential to bring to light the interplay between two interdisciplinary challenges: the scale-up of fermentation processes requiring classical engineering skills, and the engineering of microbial hosts for maximum product yields, based on comprehensive knowledge of biology. The physiological state of microorganisms and its impact on growth and product formation is the result of complex interactions between the cell and its environment. Large-scale studies have shown that homogeneous culture conditions are difficult to establish (Delvigne, Destain, et al., 2006; Lara, Galindo, et al., 2006; Takors, 2012). In the common correlation used for stirred tank reactors, mixing time τ_{mix} is proportional to $\left(PV^{-1}\right)^{-\frac{1}{3}}$, which indicates that mixing times increase as power-to-volume inputs are reduced. The latter situation is found typically in large-scale bioreactors, due to limits on power supply. This deficiency ultimately translates into regimes with extreme local differences in substrate availability, which in turn creates a demand for a high adaptation capability at the transcription level of the microorganism (Lara, Leal, et al., 2006; Löffler, Simen, Jäger, et al., 2016b; Löffler, Simen, Müller, et al., 2017; Buchholz et al., 2014; Wulffen, Ulmer, et al., 2017; Ziegler et al., 2020).

With the advent of metabolic engineering in the 1990s, the field of process optimization no longer focused solely on the extracellular environment (e.g. cultivation conditions), but started to investigate intracellular mechanisms as well (Bailey, 1991; Vallino et al., 1993). Since then, intracellular reaction rates have been quantified, and models of regulatory processes have been derived, with the ultimate aim of identifying targets for further improvements in strains and processes; these models are focus of interest (Westerhoff et al., 2004; Lee, Park, et al., 2007; Park et al., 2008; Becker, Zelder, et al., 2011; Wittmann and Lee, 2012). Many research projects have shown that cells react in a multi-response, multi-layer fashion, comprising the on- and offset of transcriptional regulation programs, as well as proteomic and metabolic changes (Teleki et al., 2015; Löffler, Simen, Jäger, et al., 2016b; Löffler, Simen, Müller, et al., 2017; Wulffen, Ulmer, et al., 2017; Buchholz et al., 2014; Oosterhuis and Kossen, 1984; Neubauer, Åhman, et al., 1995). The latter are the subject of state-of-the-art approaches that mirror the instantaneous metabolic response on extracellular heterogeneities (Wang, Haringa, et al., 2020; Haringa, Deshmukh, et al., 2017; Haringa, Tang, Wang, et al., 2018; Kuschel, Siebler, et al., 2017). To develop

gene regulatory models, comprehensive data sets are necessary; they are generated to answer the relevant biological question of interest. As noted in previous research, the consideration of transcriptional and translational effects introduces different time scales of cellular response (Löffler, Simen, Jäger, et al., 2016b; Löffler, Simen, Müller, et al., 2017; Brown et al., 2014; Wulffen, Ulmer, et al., 2017; Liu et al., 2017). Consequently, initiation and execution may be spatially disconnected, which presents a situation that differs fundamentally from many instantaneous metabolic responses studied so far (Nieß et al., 2017; Zieringer, Wild, et al., 2021). Since the quantity of gathered data is typically limited to one omics data set due to high costs, molecular translational processes can be modeled in a coarse agent-based framework solely based on transcriptomic, data which aligns very well with data from experiments, as shown in Nieß et al. (2017).

The constant stress on the microorganisms is reflected in massive product-yield losses and unwanted by-product formation (Lara, Palomares, et al., 2016; Delvigne, Destain, et al., 2006; Bylund, Collet, et al., 1998). Unfortunately, up until the present day only little is known about the bioreactor environment that the microorganisms encounter in industry-scale bioreactors (up to 100 m^3), due to the limits of current measurement techniques. Only a few experiments have been performed within the context of academic research, because of the enormous costs associated with using and maintaining large-scale equipment (Bylund, Castan, et al., 2000; Bylund, Guillard, et al., 1999; Enfors et al., 2001). In consequence, researchers have relied on the use of computational fluid dynamics (CFD) to simulate reactor flow fields, and on scale-down reactors to perform experiments for the investigation of selected scenarios (Kelly, 2008; Takors, 2012). With the use of scale-down reactors, it is possible to gather huge amount of data in order to assess the impact of cell history on the microbial and population behavior. Various designs for scale-down reactors exist, and they have been extensively reviewed (Delvigne, Takors, et al., 2017; Delvigne, Destain, et al., 2006; Neubauer and Junne, 2010; Papagianni, 2011; Lara, Palomares, et al., 2016). These setups aim to mimic simplified heterogenic environments, enabling researchers to study the consequences of insufficient mixing on microbial behavior (Löffler, Simen, Jäger, et al., 2016b; Wulffen, Ulmer, et al., 2017; Löffler, Simen, Müller, et al., 2017; Ziegler et al., 2020). However, the design of such scale-down reactors, with regard to capturing a realistic frequency and the duration of certain fluctuation events, relies in turn on particle tracking results from CFD simulations. Thus, the two disciplines are highly interdependent.

In CFD, local conditions can be predicted by solving fundamental physical equations. In the process, the fluid flow is simulated with the help of reactor-specific characteristics,

such as power input and feeding rates, which enables the prediction of mixing times and local substrate concentrations (Werner et al., 2014; Montante et al., 2005; Coroneo et al., 2011). It is only possible to biologically evaluate the identification of reactor-specific bottlenecks in substrate distribution when the frequency and duration of such fluctuations encountered by microorganisms are known. One of the tools used most often to capture these parameters in CFD is the Euler-Lagrange approach. The liquid is treated as a continuum (Eulerian representation), and the microorganisms are included as discrete massless particles following the fluid flow (Lapin, Müller, et al., 2004; Lapin, Schmid, et al., 2006; Haringa, Tang, Deshmukh, et al., 2016; Kuschel, Siebler, et al., 2017). By recording the particle trajectories, it is possible to obtain the environmental history of cells. This approach is currently limited, however, by the availability of computing power, making it necessary to use coarse simplifications.

The core engineering activities of scale-up and the analysis of bacterial behavior are both involved in a process of steady development trying to optimize the productivity in large-scale processes. That is why this thesis aims to combine both approaches for a more comprehensive point of view. CFD enables the resolution of local flow pattern and concentration profiles in bioreactors making it possible to observe particle distributions and the microbial environment over a distinct period of time (Haringa, Tang, Deshmukh, et al., 2016; Haringa, Deshmukh, et al., 2017; Lapin, Müller, et al., 2004; Kuschel, Siebler, et al., 2017). In combination with a transcriptional and translational model, based on data derived from scale-down experiments, the consequences of such concentration fluctuations (taking into account the respective duration of each individual fluctuation) can be evaluated so as to optimize the design of both.

1.2 Motivation and scope of this thesis

This thesis provides a methodology for combining and analyzing transcriptomic data derived from scale-down experiments with CFD simulation, and meets the need for knowledge-based process scale-up by elucidating the putative contributions of modeling. It will be shown that a combination of interdisciplinary tools is able to predict changing intracellular energy levels *in silico*, due to the constant adaptation of microbial producers in heterogeneous environments.

The problems highlighted above comprise the base for the research scope of this thesis. The goal was to develop methodologies to study the environment in large-scale bioreactors, and its impact on the transcriptional and translational adaptation processes of *E. coli*.

The first of these objectives was achieved by simulating a 54 m^3 reactor to identify local extracellular substrate conditions. The second objective was reached by identifying significantly perturbed genes and dynamic regulatory patterns from transcriptomic data sets of *E. coli*, to visualize ongoing adaptation processes and their impact on the demand for energy.

Research by Löffler, Simen, Jäger, et al. (2016b); Ziegler et al. (2020) and Haringa, Tang, Deshmukh, et al. (2016); Kuschel, Siebler, et al. (2017) provided the biological datasets and the implementation strategy for the Euler-Lagrange CFD approach in Ansys Fluent, respectively. The core steps required to fulfill this goal are:

- Development of a transcriptomic data analysis workflow
- Differential gene expression and gene expression dynamics analysis
- Implementation of a dynamic regulatory network of *E. coli*, including transcription and translation
- Setup of the Euler-Lagrange method in Ansys Fluent, simulating an industry-relevant case
- Combination of Lagrangian trajectory results with the gene regulatory model
- Evaluation of heterogeneity consequences, using ATP as a general biological energy currency

To extend the acquired knowledge from single to multiphase conditions, oxygen mass transfer coefficients in a 0.0015 m^3 stirred tank reactor with a gas-oil-water dispersion were investigated. In this context, the potential of safflower oil as an oxygen mass transfer enhancer was monitored, along with its impact on bubble surface tension and size distribution.

A detailed overview of the thesis outline is given in the next section.

1.3 Thesis outline

Chapter 2 includes theoretical knowledge on the biological background of *E. coli* and its physiological properties, in addition to examples of applications using oil in bioprocesses, as well as basic methods used in scale-down reactors, transcriptomic data analysis and the computational fluid dynamics used in this study. The main part of this thesis consists of peer-reviewed published papers that include the results of the studies conducted that are

described in chapters 3–5, each of which stands as an individual research project, with a description, results, discussion and conclusion. In this context, chapter 3 highlights the potential of genome analysis, based on the example of *E. coli*, by answering the following research question:

F1: How can a newly engineered strain be evaluated to trace behavioral changes back to the engineered chassis?

Transcriptomic analysis was used in chapter 4 to obtain more detailed information on microbial behavioral changes and adaptation strategies in heterogeneous environments, by answering the following question:

F2: What information is gained from differential gene expression analysis?

An example in chapter 4 of an application of transcriptomic analysis of *E. coli* SR with a modulated stringent response optimized for large-scale applications was used to identify its behavior in fluctuating nitrogen conditions, in comparison with its parental strain *E. coli* MG1655. This investigation answered the following research question:

F3: How does the engineered strain *E. coli* SR cope with repeated nitrogen starvation in comparison with its parental strain?

Subsequently, chapter 5 illustrates the formation of a glucose gradient in a 54 m^3 large-scale reactor with a combined prediction of energy levels during the adaptation process in *E. coli* MG1655, including both transcription and translation. Since transcriptomic analysis solely provides complex information about transcript levels, which are inherently coupled to protein formation, the following three research questions arose:

F4: Is it possible to extend knowledge gained through transcriptomic analysis to translation?

F5: How can transcriptomic information be translated to gain information about the overall condition of cells in relationship to production capacity?

F6: Which model is suitable for displaying these dynamic regulatory changes?

For a deeper understanding of substrate distribution in large-scale bioreactors, various CFD methods were investigated, and a link to the biological model was implemented by answering the following two research questions:

F7: Which model is suitable for displaying substrate distribution in a large-scale scenario?

F8: How can two computational intensive analysis methods be combined?

As the basis for a comprehensive, *in silico* scale-up analysis strategy, the overall goal of the research topic investigated in chapter 4 is to investigate the impact of substrate fluctuations on the non-growth-dependent ATP maintenance demand and adaptation duration as expressed in the following question:

F9: How much ATP does *E. coli* MG1655 need for starvation adaptation in a large-scale reactor and how long does it take until the adaptation process is complete?

Alongside the main topics of this thesis, summarized in research questions 1 to 9, an additional project investigating oil in water dispersions and their ability to enhance oxygen mass transfer was carried out. Preliminary investigations have lead to the following research question:

F10: How does safflower oil influence bubble surface tension, size distribution and the oxygen mass transfer coefficient?

The results of the experiments are included in Appendix E.

All of the individual research topics and questions which arose during this study are discussed and embedded within the scientific context described in chapter 5. In conclusion, this thesis presents valuable tools for use in predicting the impact of environmental stress on microorganisms and their energy availability, with attention to aspects relating to both process engineering and biology processes and characteristics.

Chapter 2

Theory

This section provides the theoretical background of methods selected in the further course of this dissertation. Starting with the biological background which introduces the effect of heterogeneous culture conditions on microbiological behavior focusing on transcription and translation processes, followed by the challenges of scale-up which promote the use of scale-down reactors, the focus is subsequently shifted towards fluid dynamic considerations to mimic realistic large-scale conditions.

2.1 *Escherichia coli*

There is no other organism so well studied and researched as *E. coli*. It is a facultative anaerobe, one of the fastest growing (growth rate $\mu = 1.73\,\mathrm{h}^{-1}$ (Cox, 2004)), rod-shaped and gram negative bacterium with a cell length of 1-5 µm. Since it was discovered in 1885 (Escherich, 1988) a large number of phenotypes in a variety of different habitats have been found. Because of this broad spectrum of lifestyles, *E. coli* is a well suited model organism to study bacterial evolution and adaptation to different growth conditions. Even if not all phenotypes are non-pathogenic (Leimbach et al., 2013), *E. coli* K-12 for example received the status as "Food and Drug Administration" (FDA) approved bacterium which paved the way as one of the most important multi-purpose bacterial cell factories in biotechnological processes today (Wittmann, Liao, et al., 2016). *E. coli* is predestined for large-scale production processes and its value to industry has even more improved since metabolic engineering was used to replace native production hosts like *Clostridium species* for production of 1-butanol (Shen et al., 2011) or 1,3-propanediol (PDO) (Tang, Tan, et al., 2009). However, the substantial value and concomitant problems of *E. coli* is best described by the biotechnological production of insulin. *E. coli* was engineered to be one of the main global producers of Insulin in the late 1970s. In 2018, it was estimated that 405.6 million people suffered from Type 2 diabetes, and this number is projected to increase to

approximately 510.8 million by the year 2030. Based on these estimates, the global usage of insulin is estimated to rise from 516.1 million vials (1000 IU) to 633.7 million vials in 2030 (Basu et al., 2019). Concomitantly, the price of one type of insulin (insulin lispro) increased 585 % (from $35 to $234 per vial) from 2001 to 2015. However, irrespective of the market price the current producers of insulin would not be able to cope with the rapid demand of affordable insulin as a result of high production costs and production capacity limitations without further development (Baeshen, Baeshen, et al., 2014).

E. coli's ability as a metabolic generalist enables the production of a large variety of compounds (amino acids, organic acids, alcohols, diols and many more (Theisen et al., 2017)). The production yield is thereby coupled to the availability of metabolites serving as product precursors and following this line of thought from energy as fuel for reactions to ensure precursor formation and transport (Zhou et al., 2009; Hara et al., 2015). Previous research suggested that, ATP pathway modifications through metabolic engineering might be an alternative way to enhance the final product biosynthesis. For instance, Zhao found that the increase of ATP supply would improve the production of terpenoid compounds (Zhao, Li, et al., 2013).

However, the opportunity to adapt the genome still incorporates the optimization under realistic process conditions and as microorganisms are developed in lab-scale the performance might be drastically hindered under large-scale conditions. Methods to identify potential sources of such deficiencies are discussed in the following chapters.

2.2 Biological Basics of Gene Expression and Protein Production

As introduced in chapter 1.1, large-scale fermentation processes struggle to provide homogeneous culture conditions for its microbial host. A heterogeneous cellular environment triggers microbial stimuli in a number of ways resulting in changes in gene expression and protein production. Such stimuli (e.g. lack of nutrient) are usually processed via a signaling cascade resulting in a changed cellular behavior. A graphical overview is given in figure 2.1. Due to its complexity, only the primary output was implemented in an agent-based model as indicated in dark red box figure 2.1. The following chapters provide the corresponding biological background and its constraints.

2.2.1 Transcription

Transcription describes the synthesis of RNA by an enzyme complex and is one of the basic processes responsible for gene expression (Watson et al., 2014). Plenty of RNA classes

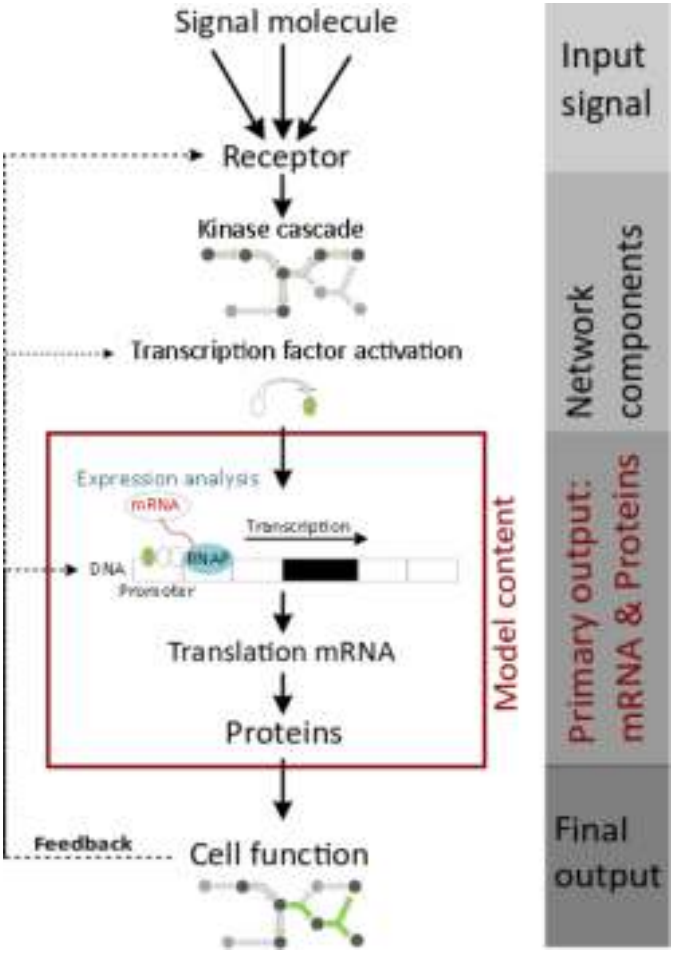


FIGURE 2.1: Workflow illustrating a signaling cascade within the cell incorporating gene expression. An extracellular signal molecule binds to a receptor at the membrane of the cell which is translated over kinase cascades into an intracellular response. Such cascades result in an activation of a transcription factor initiating (or inhibiting) the transcription of a DNA segment where the new target gene (black segment) is located. The resulting mRNA is translated into proteins which catalyze reactions and lead to new functional pathways to deal with the new extracellular changes. During the gene expression analysis, mRNA is investigated representing the expression levels of single genes. Only the primary output, as indicated in dark red, is included in the model framework.

exist in prokaryotes: rRNA, tRNA, sRNA, mRNA and many more. rRNA is part of the ribosomal complex and therefore involved in the translation process as well as tRNA, which transports amino acids to the translating ribosomal complex (Nordheim et al., 2018). Both RNA types are non-coding RNAs, thus can not be translated into proteins and do not take part in gene expression regulation (Nordheim et al., 2018). Interestingly, in the case of sRNA, another non-coding RNA, which consists of 40 - 400 nucleotides (Nt) (Storz et al., 2004), an interaction with RNA and various other regulatory processes can be observed (Vanderpool et al., 2004) indicating an important role in the gene regulatory process (Zieringer, Wild, et al., 2021). However, the regulatory function of sRNA was excluded from transcriptomic data analysis and the scope of this thesis.

Out of all existing RNA classes, mRNA is in the focus of gene expression analysis. As coding RNA, its sequence can be translated into a polypeptide strand. The transcription process is composed of several steps, which are grouped in three phases: initiation, elongation and termination (Watson et al., 2014).

Initiation

The initiation by RNA Polymerase (RNAP) takes place at a unique position upstream of the coding sequence (Lodish, 2000), called promoter (figure 2.1). Several different σ-factors exist in *E. coli* which direct the RNAP complex to different promoters (Helmann et al., 1988; Watson et al., 2014). The rate by which a transcription process is initiated is a measure for the strength of the corresponding promoter sequence and is not only influenced by the binding of RNAP, but also by the speed of melting of the DNA double strand and promoter escape (Rhodius et al., 2010). Additionally, there are regulatory proteins called activator or inhibitor, that enable or suppress the binding of RNAP. Mostly, they bind within or upstream of the promoter sequence and appear to act directly with RNAP (Busby, 1994; figure 2.1). Activators and inhibitors are also called transcription factors and control the transcription process. As they are encoded on the DNA and are subject to regulation, a complex regulatory network is formed (Machado et al., 2011). After binding to the promoter region, the RNAP complex needs to escape from this region to terminate initiation and start the elongation process. This promoter escape could be a rate limiting step for many promoters in the transition from initiation to the elongation phase (Reppas et al., 2006). After promoter escaping and synthesis of a few nucleotides by RNAP, the σ-factor dissociates from the core enzyme and terminates the initiation phase (Gross et al.,

1998). In the context of the agent-based model, the initiation of transcription is inherently included in the transcriptomic dynamics derived but no individual information can be assessed about promoter strength or other preceding interactions.

Elongation

During the elongation phase, the RNA polymer is synthesized in a repetition of three reaction steps, as described by Nordheim et al. (2018):

- Binding of a suitable nucleotide triphosphate (NTP) at the active center of the enzyme
- Release of a pyrophosphate molecule and generation of a phosphodiester bond between the nucleotide and the 3'-OH-cap of the RNA strand
- Movement of the RNAP relative to the DNA strand

It is likely, that these steps do not proceed with constant velocity, instead they are accompanied by delays or even stops of the RNAP movement, due to falsely attached nucleotides or DNA bound proteins, that sterically hinder RNAPs (Nordheim et al., 2018). This pausing is a regulatory effect by which the cell can synchronize transcription and translation or by which a time window is created that allows the interaction of further regulatory proteins (Artsimovitch et al., 2000). It is even possible, that the arrested RNAP encountered by a protein dissociates from the DNA eventually and the elongation is terminated (Watson et al., 2014). Different values are reported for the average elongation speed. Bremer et al. (1996b) report values around 50 Nts^{-1}, while Chen, Shiroguchi, et al. (2015b) found a value of 25 Nts^{-1} for exponentially growing *E. coli* cells. Furthermore, they found a reduction in the average elongation rate in the stationary phase with a decrease to 21 Nts^{-1}. Tolić-Nørrelykke et al. (2004) reported varying constant elongation velocities at different loci on the DNA, dependent on the transcription elongation complexes that are formed by different RNAP which is potentially caused by post translational modifications of RNAP. Apparently, this rate is not uniform but the exact mechanisms remain unknown. That is why an averaged constant elongation rate was considered for all moving RNAPs in the model framework (see chapter 5). The energy for synthesizing a mRNA strand is delivered by the cleavage of two phosphate groups from one NTP. After degradation of monophosphorylated nucleotides, recycling requires two ATP per molecule in

return (Löffler, Simen, Jäger, et al., 2016a). *De novo* synthesis of nucleotides instead requires more energy and differs for each nucleotide type as different precursors are needed for production (see chapter 5).

Termination

After reaching a terminator codon, the RNAP separates from the template. This can either happen rho-dependent or independent with a protein called rho factor or a stable loop sequence in the DNA causing the polymerase to stall. After separation RNAP returns to its pool or immediately initiates transcription again. The resulting mRNA is released (Peters et al., 2011).

2.2.2 Translation

Translation describes the next step in protein synthesis and considers the proceeding from a mRNA molecule to the final protein. As transcription, translation is divided in three steps: initiation, elongation and termination (Watson et al., 2014).

Initiation

Initiation in general is considered to be the most relevant rate-limiting step in mRNA translation with a variety of proteins involved (McCarthy et al., 1990). The translation starts at a certain start codon which mostly (around 90%) consists of AUG, and with around 8% less frequently of GUG (McCarthy et al., 1990). The translation initiation sequence which is located at the 5′-end of the start codon is called Shine-Dalgarno sequence (SDS) and is complementary to a sequence of the 16S ribosomal subunit (Watson et al., 2014). The extent of complementarity of the SDS and its anti-sequence as well as the spacer between the SDS region and the start codon strongly influence the frequency of initiation of a certain mRNA (Watson et al., 2014). However, the SDS is neither necessary nor sufficient for initiation as there exist a variety of SDS-lacking mRNAs that are still translated (Gualerzi et al., 2015).

Only a simplified mechanism is considered for initiation of translation in the model incorporating a minimum distance between moving molecules on the mRNA strand as

threshold. As soon as a minimum distance between RNAPs is reached, ribosomes are able to bind and translate the mRNA sequence in a co-transcriptional manner.

Elongation

After the ribosome locates on the mRNA elongation starts (Subramaniam et al., 2014). In general, it consists of three steps, which are repeated until a stop codon is reached. According to Wintermeyer et al. (2004), in the first step a aminoacyl tRNA (aa tRNA), carrying an amino acid (AA), is transported to the ribosome. This step is facilitated by a protein called elongation factor Tu and requires the energy of GTP dephosphorylation. The structure of a tRNA consists of a certain anti-codon, which corresponds to a complementary codon on the mRNA (Watson et al., 2014). Consequently, a variety of differently loaded tRNA coexist and the anti-codon sequence ensures, that the suiting one is bound to the polypeptide. Next, the aa tRNA forms a peptide bond at the carboxyl end of the peptidyl tRNA, carrying the peptide chain. The formation of a peptide bond does not require hydrolysis of triphosphate molecules, instead the reaction is driven by breaking the high-energy acyl bond between the tRNA and its amino acid. This reaction and therefore the loading of a tRNA requires two-fold hydrolysis of an ATP molecule to adenosine monophosphate (AMP) (Watson et al., 2014). Afterwards, the peptide is extended by one AA while a mere tRNA, without any loaded AA leaves the ribosomal complex. Finally, the ribosomal complex steps one codon forward and a free position for a new cycle remains. This step is facilitated by a protein called EF-G which requires the cleavage of a guanosine triphoshate (GTP) molecule. The speed of elongation is about 20 AA s^{-1} with an error rate of 0.5 - 0.1 ‰ (Nordheim et al., 2018).

Termination

Termination of translation is indicated by three possible stop codons: UAG, UAA and UGA (Nordheim et al., 2018). As cells normally do not contain tRNA with a complementary sequence to one of these codons, the ribosomal complex stops when it reaches a certain codon. As soon as a stop codon is reached the complex disrupts and releases the mRNA strand as well as the peptide strand. The ribosomal units are then recycled to initiate a new cycle of translation.

2.2.3 Transcript Degradation

The active degradation of mRNA is a central function of the cell's regulatory machinery (Belasco et al., 1988). Thereby, life time varies between 30 s and several minutes. On average, life time of transcripts was estimated to be 2.5 min for exponentially growing *E. coli* cells, while an increase in transcript stability is observed for cells in stationary phase. Here, an average value of 4.5 min for the life time of transcripts is observed (Chen, Shiroguchi, et al., 2015b). The process of mRNA degradation involves several endo- and exonucleases (Hui et al., 2014). Exonucleases use either hydrolysis or phosphorylation to cleave a nucleotide from the mRNA strand. While hydrolysis irreversibly results in a monophosphorylated nucleotide, the phosphorylation reaction is reversible and produces a nucleotide diphosphate (Hui et al., 2014). By starting degradation of a transcript at its 5′-cap, a cell can ensure, that initiation of translation is inhibited by degrading the SDS for ribosomal binding and already initiated translation can successfully be terminated (Laalami et al., 2014). Contrary to degradation from the ends, endonucleases break bonds within a mRNA molecule. Resulting fragments after cleavage are then degraded by exonucleases. In contrast to initiation of transcription and translation, the initiation event of endonucleolytic cleavage cannot be categorized by a specific interaction (Carpousis et al., 2009). mRNA decay in *E. coli* is typically initiated by endonucleolytic cleavage followed by an exonucleolytic degradation of the fragments (Viegas et al., 2011). The rate of degradation is influenced by many parameters like the abundance of nucleases, polyadenylation or the 5′-phosphorylation status (Silva et al., 2011). However, one of the most important factors preventing mRNA strands from degradation is its loading with ribosomes (Laalami et al., 2014), which was set as constraint in the biological model.

2.2.4 Protein Degradation

Protein degradation is an active and important process for cell viability (Maurizi, 1992a). Due to the irreversible nature of proteolysis, cells have to take care of arbitrary degradation and therefore ensure tight regulation (Mahmoud et al., 2018). In general the turnover rates of different proteins varies. The major fraction of the total protein pool shows lifetimes that exceed, sometimes by far, the doubling time of cells. Short lived proteins account for up to 40 % of all synthesized proteins and due to the great number it is unlikely that they are all accidentally or falsely synthesized proteins, but instead fulfill e.g. regulatory functions (Goldberg et al., 1976). Generally proteolysis is divided into housekeeping

degradation and a regulated part (Maurizi, 1992a). Main issue of the housekeeping degradation machinery is prevention from damage by non-functional proteins, that were falsely transcribed or translated, experienced chemical or oxidative damage, improper folding or thermally induced structural changes (Maurizi, 1992a). Concerning adaptation to environmental changes, regulated protein degradation is a key step to ensure persistent rapid protein activation (Gur et al., 2011). This degradation is carried out by energy-dependent enzymes (Mahmoud et al., 2018).

Although proteolysis requires ATP, regulation at the protein level can exhibit advantages that outweigh the energy cost, e.g. a regulator protein can be deactivated rapidly which might be useful in response to heat shock or general stress response (Jenal et al., 2003). Moreover it is observed, that protein turnover is altered during nutrient depletion and proteolytic rate is increased (Pine, 1973). This might help to supply the cell with amino acids for renewed protein synthesis (Maurizi, 1992a), which was considered as recycling in the model.

2.3 Data Acquisition and Analysis Methods

Signals like environmental changes feed into the transcriptional regulatory systems, which affect the physiological and morphological changes that enable organisms to adapt effectively for survival (Kitano, 2000). For instance, Ishii et al. (2007) generated a dataset for systematic analysis of *E. coli* cells under genetic and environmental perturbations, showing that the metabolism responded highly flexible. Observations like this build the basis of the following chapters, where experimental setup and computational methods are introduced to unravel these kind of impairments and offer an inexpensive routine, rather than requiring significant production-scale efforts.

Experimental Set-ups mimicking Large-scale Heterogeneities

Scale-up is the procedure that transfers lab-scale bioprocesses in production scale, often covering 7 to 8 volume of magnitude, aiming to achieve larger product quantities, with simultaneous increase or at least maintenance of yields and product quality. Unfortunately, loss or even failure of large-scale performance may occur often caused by micro-environmental inhomogeneities. Insufficient mixing leads to severe axial and horizontal concentration gradients. Producer cells frequently cross these poorly mixed zones which trigger metabolic and transcriptional responses accordingly (Takors, 2012) redirecting the

precursor and energy supply from production to adaption (Zieringer, Wild, et al., 2021). Thus, detailed knowhow is necessary to prevent non-wanted production losses.

While geometric properties of the reactor can be maintained over different scales, accordance of chemical and physical quantities are much harder to achieve. One of the most deteriorated parameters throughout scale-up is power input (Oldshue, 1966; Takors, 2012). As in lab-scale (< 0.05 m^3) ideally mixed systems are common it can be hardly achieved in large-scale (> 10m^3) (Garcia-Ochoa et al., 2009; Vrabel et al., 2000). If the volumetric power input is kept constant, as it is one of the rules for scale-up in classic process engineering, high mixing times and large substrate gradients might occur.

Because of a lack of large-scale experimental data due to their immense costs and capacity needs, scale-down device experiments are a cost-reducing alternative and resemble large-scale conditions very well (Delafosse et al., 2014; Takors, 2012). Accordingly, Oosterhuis and Kossen were the first who presented a scale-up simulator (1983) comprising two stirred tank reactors (STRs) for investigating the impact of oxygen gradients on *Gluconobacter oxydans* (Oosterhuis, Groesbeek, et al., 1983). They further introduced bioreactor compartment models to achieve coarse spatial resolution of local oxygen transfer rates to identify micro- and anaerobic zones (Oosterhuis and Kossen, 1984). This line of thinking was followed by a series of likewise studies (Neubauer, Häggström, et al., 1995; Buchholz et al., 2014; Löffler, Simen, Jäger, et al., 2016b; Löffler, Simen, Müller, et al., 2017; Wulffen, Ulmer, et al., 2017). Reviews have been given by Delvigne *et al.* and Neubauer *et al.* (Delvigne, Takors, et al., 2017; Delvigne, Destain, et al., 2006; Neubauer and Junne, 2010). Within this variety of scale-down devices the STR-PFR cascade is highlighted in many comparative studies (Delvigne, Takors, et al., 2017; Delvigne, Destain, et al., 2006; Neubauer and Junne, 2010) and is successfully established as standard scale-down device for the simulation of large-scale conditions (Limberg et al., 2016). Thereby, the STR is operated as a well-mixed compartment under standard limited growth conditions and the plug flow reactor (PFR) simulates a feeding, starvation or anaerobic zone providing the stimulus to be investigated (Lara, Leal, et al., 2006). The latest approaches used an improved revision of a STR-PFR device, recently published by Löffler, Simen, Jäger, et al. (2016b) and Wulffen, Ulmer, et al. (2017). In this setup a distinct reference state sample port was established in the STR. Hence, they were able to detect short term adaptation of the cells to fluctuating conditions and long term changes initiated after a longer time span as well as the formation of a new steady-state in comparison to the reference state. Starvation zones have attracted attention as relatively huge zones of poor substrate availability were identified in large-scale reactors (Neubauer, Häggström, et al., 1995; Lapin, Müller,

et al., 2004; Haringa, Tang, Deshmukh, et al., 2016). From CFD simulation and measured data it is known that distant from the feeding point or close to the reactor walls poorly mixed zones with very low nutrient concentrations exist. According to findings of Lapin, Schmid, et al. (2006), who investigated the growth performance of *E. coli in silico* under well and poorly mixed fed-batch conditions, the volume ratio PFR-to-STR was designed. Since then, other researchers made use of CFD to further optimize or develop scale-down reactors (Kuschel and Takors, 2020; Haringa, Deshmukh, et al., 2017). Successful scale-down of heterogenic large-scale bioreactor environments by rules established by classic process engineers is only the first step in unraveling the impairments on the microorganism. The next sections introduce the access to biological responses triggered by nutrient starvation.

2.4 Access to Biological Response in Starvation Conditions

With a scale-down reactor, microbial behavior can be investigated under different large-scale scenarios. Samples taken from the scale-up simulator need to be processed so that the cellular state is frozen immediately. For genomic analysis biosuspension is usually sampled, flash-frozen in liquid nitrogen and stored at -70 °C. On the day of extraction samples are thawn and total DNA extracted with a corresponding commercial kit (e.g. DNeasy Blood and Tissue Kit (Qiagen)). Blocking intracellular transcription is achieved by sampling into RNA protect kits (e.g. RNeasy Protect Bacteria Kit (Qiagen)). Afterwards, samples can be treated to identify nucleotide and transcript levels by either applying microarrays or, more preferred, next generation sequencing (NGS) technologies analyzing DNA or mRNAs like DNA- or RNA-Seq (Shendure et al., 2008; Wang, Gerstein, et al., 2009). In preparation for sequencing almost all steps of the various protocols have been reported to introduce bias, especially in the case of RNA-Seq, which is technically more challenging than DNA-Seq.

Knowledge of the nature of these biases will be essential for a careful interpretation of NGS data and methods to check for certain biases is given in the next chapters. A detailed review on this topic is given by Van Dijk et al. (2014).

2.4.1 Genomic DNA Sequencing

Since the early 1990s, DNA sequence production has almost exclusively been carried out with capillary-based, semi-automated implementations of the Sanger biochemistry (Hunkapiller et al., 1991). The starting material for DNA-Seq is generally double-stranded

DNA in the form of isolated genomic DNA. This DNA is fragmented, followed by end-repair and adapter ligation, and usually a size selection step to remove free adapters and to select molecules in the desired size range. Next, polymerase chain reaction (PCR) amplification is often performed to generate sufficient quantities of template DNA to allow accurate fragment (also called read) quantification (Linnarsson, 2010). Sequence is determined by high-resolution electrophoretic separation of the single-stranded, end-labeled extension products in a capillary-based polymer gel. Laser excitation of fluorescent labels as fragments of discreet lengths exit the capillary and provides readout that is represented in a Sanger sequencing 'trace'. Software translates these traces into DNA sequence, while concomitantly generating error probabilities for each base-call (Ewing and Green, 1998; Ewing, Hillier, et al., 1998). To reduce the risk of biases accompanying these preparation steps (Van Dijk et al., 2014), specific parameters can be checked for consistency and thresholds: GC content, quality score Q, mappability of sequencing reads and regional biases that might be generated by local structure which is explained further in the following sections.

Whole-genome sequencing can detect single nucleotide variants, insertions/ deletions, copy number changes and large structural variants. It provides a high- resolution, base-by-base view of the genome, identifies potential causative variants for further follow-up studies of gene expression and regulation mechanisms and delivers large volumes of data in a short amount of time to support assembly of novel genomes (Shendure et al., 2008; Costessi et al., 2018).

After three decades of technological development, the Sanger biochemistry can be applied to achieve read-lengths of up to around 1,000 bp, and per-base 'raw' accuracies as high as 99.999 %. In the context of high-throughput shotgun genomic sequencing, Sanger sequencing costs are in the range of $23 USD per sample (https:// www.illumina.com; Date: 22.01.21). Based on these values, sequencing represents a less expensive and convenient analysis method and it is going to become an individualized, universally accessible technique in the near future. This means, that every process can be preceded by a sequencing step to individually check for the microbiological needs and requirements and possible engineering targets to improve space-time-yield. In this thesis, DNA sequencing was used to determine the selective pressure on global regulatory programs (high pressure: many variants) and to validate the biological behavior due to engineered changes unlike random mutations triggered by stress.

2.4.2 RNA-Seq

RNA-Seq is a high throughput experimental technique that allows sequencing of complementary DNA (cDNA) at very high redundancy (also called depth) with up to 10^7 individual sequences (reads) per sample (Wang, Gerstein, et al., 2009). This technique can be used to examine alternatively spliced transcripts, post-transcriptio- nal modifications, changes in gene expression, and more. Unlike microarrays, RNA-Seq does not depend on prior knowledge of the genome sequence. Therefore, any preconceived notions about what to detect (via probes or primers) can be avoided, resulting in decreased overall bias. Additionally, RNA-Seq results show single-base resolution to identify Single Nucleotide Polymorphisms (SNPs) with high sensitivity. The sequencing process typically begins with RNA purification and quality assessment. Depending on the target RNA of interest (mRNA, tRNA, sRNA and many more) commercial kits for purification exists and need to be chosen accordingly. Once the RNA is extracted the purity and integrity of the samples needs to be assessed. The assessment for contaminants is done via UV spectrometry. The peak of absorbency for RNA molecules is at 260 nm, while the peak for contaminants is usually distinct (Cellerino et al., 2018). The RNA integrity is quantified in a score called RNA Integrity number (RIN) whose major proportion is determined by the shape of the peaks of the 16S and 23S rRNAs (85 % of the total RNA fraction (Bremer et al., 1996a)). The RIN value comprises an interval of 1 to 10, where 10 implies perfectly intact and 1 extremely degraded RNA due to the presence of active RNAses. To ensure high quality samples the RIN should be higher than 7 (Schroeder et al., 2006). The next step is library preparation and it usually begins with fragmentation followed by generation of cDNA via reverse transcription. Unfortunately, assembling cDNA requires additional steps leading to increased signal degradation and increased chances of sample contamination. Moreover, linker ligation for cDNA synthesis introduces bias because the linker does not ligate to different RNA end sequences with the same efficiency. cDNA synthesis also favors small and medium RNAs to the detriment of longer sequences. rRNA accounts for more than 85 % of the prokaryotic cellular RNA content (Karpinets et al., 2006), which can impede the analysis of mRNA transcripts, with library cDNAs mapping to rRNA in the absence of selection procedures (Van Vliet, 2010). Since mRNA (which is usually the RNA of interest) is the least represented in the total RNA pool (around 3 %), the most abundant species need to be removed. Therefore, approaches to address this issue have focused on removing the prokaryotic rRNAs prior to construction of cDNA libraries. All

these steps are necessary prior to sequencing. Illumina sequencing incorporates fluorescent nucleotide analogues into the cDNA at each sequencing cycle and to scan them in order to determine the sequence of the transcript (Cellerino et al., 2018). The choice between single- or paired- end sequencing depends on the experimental design and the biological target. Paired-end sequencing is more cost and time intensive than single-end sequencing (every fragment is sequenced twice), but reveals more structural information (e.g for identification of SNPs or if no reference genome is available). Sequencing also incorporates the information about the desirable read length. Usually it is in the range of 50-300 bp (Cellerino et al., 2018) with the rule: the longer the reads the higher the structural information. All these steps result in raw data requiring further processing. Various methods for RNA-Seq analysis are available and have been reviewed by Conesa et al. (2016). However, analysis results are highly platform-specific (Lam et al., 2012), which is a common problem during sequencing. Results are dependent on a standard processing and analysis workflow which does not exit yet. In the next chapters the methods and algorithms used in this thesis are introduced.

2.4.3 Assessing Quality of Sequenced Data Sets

Typically, samples are transferred to commercial sequencing partners (either DNA or RNA) resulting in fast and high quality reads (depending on the provided sample quality). Data is usually delivered as fastqc files with corresponding quality report. This report includes several different properties ranging from basic statistics (e.g. GC-content for *E. coli* around 51% (Lee, Barber, et al., 2020)) to per base sequencing quality of each individual sample. Thus, the reliability of the data set can be evaluated with the use of a quality score Q as it is described in the following and subsequent steps can be defined.

In high-throughput data analysis quality is a fundamental issue and needs to be assessed over the entire analysis. Raw data from a sequencer come in the fasta or fastq (extension of the fasta format) format and are associated to the Phred Quality score (Q). Applied to the Sanger sequencing (currently adopted by Illumina after v1.8) it is defined as:

$$Q = -10\, log_{10}(P) \tag{2.1}$$

where P is the probability that base-calling for a given nucleotide sequence is inaccurate. Reasonable base calls are considered to be above 28 for good quality bases, which equals a confidence of at least 99.8 %. Poor quality bases are Q < 20 (confidence is less than

99 %). However, this threshold highly depends on the application (e.g. lower Phred scores are acceptable if expression values from RNA-Seq data are obtained) (Cellerino et al., 2018). Additionally, as mentioned before, GC content and mappability of sequenced reads are other quality parameters. There are certain methods to improve the quality of the sequenced dataset. In the workflow established trimming was used to cut-off the low-quality parts from the reads in order to achieve higher accuracy during the mapping step to the reference genome. As soon as a sufficient phred score is achieved, quantification of gene-level expression allow the insight in cellular processes behind a certain behavior of the microbial organism.

2.4.4 Genomic Data Analysis

Before mapping, DNA reads need to be assembled by physically linking DNA fragments together creating 'contigs', a series of overlapping DNA sequences used to make a physical map that reconstructs the original DNA sequence of a chromosome or a region of a chromosome. One often used DNA assembler tool specifically designed to assemble bacterial genomes combining the accuracy of short reads and the structural resolving power of long reads is Unicycler (Wick et al., 2017). The obtained contigs need to be reordered and mapped to a reference genome where the mappability is an indicator of sample quality. Mauve represents a powerful tool to describe a genome comparison method that identifies conserved genomic regions, rearrangements and inversions in conserved regions and the exact sequence breakpoints of such rearrangements across multiple genomes. Furthermore, this method performs traditional multiple alignment of conserved regions to identify nucleotide substitutions and small insertions and deletions (indels) (Darling et al., 2004). As this method provides exactly the amount of detail to evaluate the selective pressure in form of indels or small nucleotide polymorphisms (SNPs) Mauve was integrated in the analysis workflow used in this study. However, with an error rate of over 4 % during short read sequencing, most accurate biological evaluation of possible regional biases and variants is still achieved by manual review of identified changes which concludes DNA-Seq analysis in this study (Reumers et al., 2012).

2.4.5 Transcriptomic Data Analysis with DeSeq2

Organisms can adapt to changing environments due to a flexible gene expression program controlled by the dynamic interactions of hundreds of transcriptional regulators.

To unravel this regulatory complexity, multiple computational algorithms have been developed to analyze gene expression profiles and detect dependencies among genes over different conditions. Figure 2.2 provides an overview of a typical workflow making use of public R packages.

However, experimental validation of the precision of these methods at genome scale has remained unsatisfied due to the lack of a model organism with both, a known regulatory structure and compatible experimental data. Recent development of Love et al. (2014), lead to an improved DeSeq Algorithm to unravel transcriptomic events called DeSeq2. In the next chapter the mathematical background of this algorithm and methods used in the developed workflow are introduced. In this thesis the workflow established starts from fastq files. These files were aligned to the reference genome and a count matrix which tallies the number of RNA-Seq reads within each gene for each sample was prepared. Differential gene expression analysis was performed with the Bioconductor package DeSeq2 and the results were processed graphically. RNA-Seq is not only technically more challenging than DNA-Seq but also its analysis is more complex. The following sections comprise the steps applied during RNA-Seq analysis, after quality assessment of the sequenced data set.

Read Counting

Quantification of gene-level expression allows the insight in cellular processes behind a certain behavior of the microbial organism. Currently, one of the most widely used approaches is to identify the genomic locations of a set of non-overlapping features (genes) while using the number of aligned reads as a measure of its abundance. So far there are two major feature counting tools: featureCounts (Liao et al., 2014) and HTSeq-count (Anders and Huber, 2010). Both are very well known, reliable and have their own advantages (Liao et al., 2014). Due to the good compatibility of the results from HTSeq-count to differential expression tools (e.g DeSeq2) it was integrated in the workflow of this thesis. To keep the number of potential miscounted reads (artifacts) at a minimum, it is recommended to filter data for low count genes.

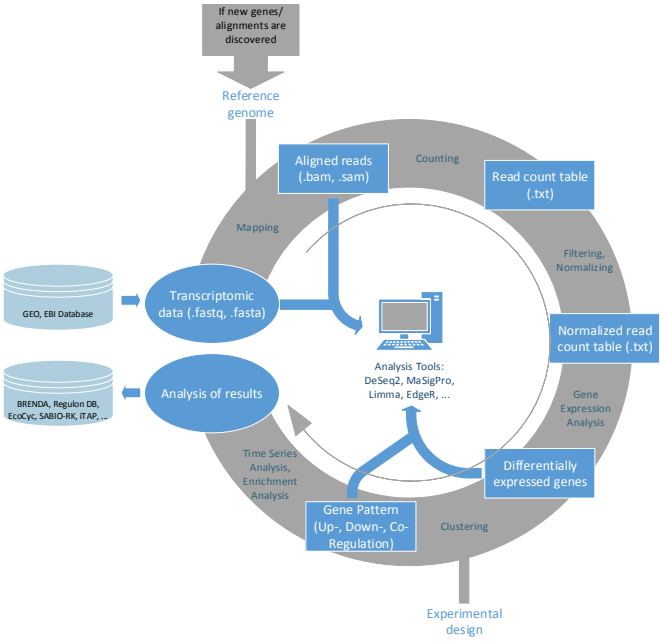


FIGURE 2.2: Workflow illustrating the general procedure when analyzing gene expression data. RNA-Seq analysis creates fasta and fastq as raw data formats compiled in gene expression repositories (GEO, EBI), followed by SAM or BAM files for aligned reads. The analysis steps are, in general: (1) Mapping the reference genome onto the transcriptomic data, (2) Counting reads, (3) Filtering low read counts and normalizing counts, (4) Gene Expression Analysis, (5) Clustering, (6) Time Series Analysis and Enrichment Analysis. DeSeq2, MaSigPro, limma and edgeR are often applied packages programmed with the language R to analyze transcriptomic data, in case of differential gene expression as well as gene pattern analysis. The resulting dynamic expressions and parameters are stored in databases like BRENDA (Scheer et al., 2010), Regulon DB (Salgado et al., 2004), EcoCyc (Keseler et al., 2013), SABIO-RK (Wittig et al., 2011) and iTAP (Sundararaman et al., 2015).

Detection and Analysis of Differentially Expressed Genes with DeSeq2

With a quality-curated raw read count table the platform for further analysis has been created. However, sound comparison of gene expression levels between different experiments can only be accomplished when transcript abundance is normalized across samples. Transcripts have different lengths and libraries originating from different samples with different sequencing depths (range of total numbers of reads). The most commonly used procedures to correct for these technical biases are: Transcripts per Million (TPM) (Li, Ruotti, et al., 2010), Trimmed mean of M-values (TMM) (used by edgeR (Robinson, McCarthy, et al., 2010)) or relative log expression (RLE) (used by DeSeq2 (Love et al., 2014)). In general, two different comparison approaches need to be distinguished, intra- and intersample comparisons. Notably, TMM and RLE do not correct the observed read counts for gene length. Therefore, both approaches do not allow for intrasample comparison, because longer genes will get more read counts compared to shorter genes when expressed at equal levels. Thus, samples can seem correlated based on gene length instead of the expression levels without additional correction. The most commonly used normalization method that includes gene length correction (as well as sequencing read length correction) is TPM (Li, Ruotti, et al., 2010) which proved to be inadequate and biased (e.g. by sequencing depth) for intersample comparison of transcript levels when used on its own (Abbas-Aghababazadeh et al., 2018; Conesa et al., 2016). These biases can be addressed in conjunction with normalization techniques such as TMM. However, in this project differential expression level analysis needs to be simple yet still appropriate for intersample comparison. Since tools for differential expression analysis are comparing the counts between sample groups for the same gene, gene length does not need to be accounted for. However, sequencing depth and library composition (e.g. if a large number of genes are unique to, or highly expressed in, one experimental condition, the remaining genes in that sample occur diminished) do need to be taken into account. That is why RLE was used to analyze the transcriptomic data in this study. RLE is a robust normalization method, e.g. produces homogeneous sets of transcript abundance (Love et al., 2014). To normalize for sequencing depth (library size) and library composition, DESeq2 uses the median of ratios method. First, the logarithm is taken of each value and all genes which are only transcribed in one sample will be removed (which accounts for different library compositions). DESeq2 then creates a pseudo-reference sample for each gene by calculating the geometric mean across all samples and divides each value by this reference. Afterwards the median of the ratios for each sample is calculated and serves

as a scaling factor to normalize the raw counts. Since the majority of genes are not differentially expressed (DE), the majority of genes in each sample should have similar ratios within the sample.

To decide whether a difference occurred randomly the underlying statistics of the dataset need to be considered. The count data generated by RNA-Seq exhibits overdispersion (variance > mean) and the statistical distribution used to model the counts needs to account for this overdispersion. Love et al. (2014) use a negative binomial model to describe the data distribution in their widely used DeSeq2 package. They start from two assumptions: (i) the majority of genes are not differentially expressed (working hypothesis H), (ii) the relationship between mean and variance is a smooth function (continuous rise of variance with increasing mean). This function is derived by performing a local regression (e.g. Poisson regression) where mean and variance of count variables are assumed to be equal or the negative binomial regression where mean and variance vary. In contrast, EdgeR models the negative binomial by multiplying the variance with a constant, instead of a local regression (Robinson, McCarthy, et al., 2010), which leads to overestimation of the mean/variance dependence at high expression values and underestimation at low expression values.

To determine if a gene is differentially expressed the probability of accepting the null hypothesis (H_0) needs to be calculated, which is the negation of the working hypothesis (H). For instance, a statistical t-test is calculated for each gene that is associated with the probability distribution of the null hypothesis. With DESeq2, the Wald test is commonly used for hypothesis testing when comparing two groups (besides likelihood ratio). A Wald test statistics is computed along with a probability that a test statistics at least as extreme as the observed value was selected at random. This probability is called the p-value of the test. If the p-value is small the null hypothesis is rejected (e.g. the gene is differentially expressed).

If the p-value is used directly from the Wald test with a significance cut-off of $p < 0.01$ (usually a reliable p-value is $p < 0.05$), that means there is a 1 % chance it is a false positive. Each p-value is the result of a single test. The more genes are tested, the more the false positive rate is inflated. This is called the multiple testing problem. For example, if 4,000 genes are available for differential expression, at $p < 0.01$ we would expect to find 40 genes by chance. If 400 genes are found to be differentially expressed in total, roughly every tenth gene is false positive. To correct for multiple testing and reduce the number of false positives one common approach is proposed by Benjamini and Hochberg (1995).

They defined the concept of False Discovery Rate (FDR) and created an algorithm to control the expected FDR below a specified level given a list of independent p-values. An interpretation of the Benjamini-Hochberg (BH) method for controlling the FDR is implemented in DESeq2 in which the genes are ranked by p-value, then multiply each ranked p-value by its rank divided by the total number of tests multiplied with a manually chosen false discovery rate (usually below 0.05).

Many common statistical methods for exploratory analysis of multidimensional data, especially methods for clustering and ordination (e.g., principal-component analysis), work best for (at least approximately) homoscedastic data. This means that the variance of an observed quantity (e.g. expression strength of a gene) does not depend on the mean and is equal among several groups. In RNA-Seq data however, variance grows with the mean. For instance, if principal components analysis (PCA) is performed directly on a matrix of normalized read counts, the result typically depends only on the few most strongly expressed genes because they show the largest absolute differences between samples. A simple and often used strategy to avoid this is to take the regularized-logarithm (rlog) of the normalized count values to form approximately homoscedastic data. For genes with high counts, the rlog transformation differs not much from an ordinary $\log_2$ transformation. For genes with lower counts, however, the values are shrunken towards the genes' averages across all samples.

Identifying Pattern in DEG Datasets

After finishing the DEG analysis it is important to extract the relevant information and conclude biological sense from it. Clustering and PCA are two widely-used approaches to reduce complexity and detect pattern within the data.

PCA is a technique for reducing the dimensionality of such datasets, increasing interpretability but at the same time minimizing information loss. It does so by creating new uncorrelated variables that successively maximize variance. Finding such new variables, the principal components, reduces to solving an eigenvalue/eigenvector problem. The new variables are defined by the dataset at hand, not *a priori*, what makes PCA an adaptive data analysis technique. Additionally, variants of the technique have been developed that are tailored to various different data types and structures (Jolliffe et al., 2016). However, principal components are not always easy to interpret and even if most information can be preserved, the principal components only allow a limited insight into the details of the underlying data.

For **clustering** a huge variety of different algorithms exits. Clustering in general describes the classification of data into groups with similar objects, where similarity in expression behavior (concerning genes) implies a relationship in biological function (Jain, Murty, et al., 1999). The concrete definition of a cluster hereby is influenced by the operator, his intentions and the operating field (Aghabozorgi et al., 2015). There is neither a standard algorithm, which generally delivers better results than another, nor is there a uniform opinion which criteria are essential and significant for cluster quality (D'haeseleer, 2005). In this thesis we will focus on partitioning and quality-based clustering and introduce the application options in the following. A general advantage of the two used algorithms compared to hierarchical clustering is that clusters are generated in an iterative process, in which the single members are swapped until an ideal composition has been found (Backhaus et al., 2018; Theresa Scharl et al., 2006). Hierarchical algorithms do not exhibit this property, instead they proceed either in a step-wise agglomerative combination of smaller clusters to one larger clusters or, in a diversive approach, in which a large group is further divided into smaller clusters (Halkidi et al., 2001). According to Drăghici (2012) and Nagpal et al. (2013), one main disadvantage of hierarchical clustering is that important division steps are executed at the beginning of the sorting process, at a point in time where only little is known about the actual structure of the data. Also, one challenging task is to decide when a new cluster is formed (Moreau et al., 2002) with no biological reason that justifies the strict binary splitting of a cluster (van der Laan, Mark J. et al., 2003) as e.g. multiple genes can be controlled by the same operator.

K-Means Clustering: The K-means algorithm is one of the most used techniques, because it is algorithmically simple, relatively robust and gives 'good enough' answers over a wide variety of data sets (Lloyd, 1982; Drăghici, 2012). It belongs to the class of partitioning methods often using euclidean distance as similarity measure. In the beginning, a fixed number of clusters needs to be specified. According to Jain (2010), centers are randomly chosen for all predefined clusters at the beginning and the data points are joined with the center closest to them. Next, new cluster centers are calculated based on the new members that were added and new data points are assigned to the center closest to it. This is repeated until clusters' compositions are fixed. K-means clustering has the advantage of an easy implementation and fast computing. One noteworthy disadvantage of this procedure is the *a priori* specification of the total number of clusters, as they are often unknown (Chandrasekhar et al., 2011). Furthermore, the algorithm is not deterministic due to the randomly chosen start centers and the resulting convergence to local optima

(Lu et al., 2004). This leads to different results for different cycles. Additionally, positioning of clusters can be misleading, as two points that are close to each other, can still be sorted into different clusters and are a result of the position of the initial cluster centers. A general problem, that inhibits K-means, is the estimation of the number of clusters (Zhao and Karypis, 2005). Furthermore each gene is forced to be part of a cluster, regardless of its relevance (Moreau et al., 2002). This makes the algorithms vulnerable to outliers and noisy measurements. Addressing these intrinsic problems of the „traditional algorithms“ (Oyelade et al., 2016), methods like quality-based clustering (Heyer, 1999) were introduced, invented specifically for clustering gene expression data (Jiang et al., 2004; Moreau et al., 2002).

Stochastic Quality-based clustering (SQBC) is an advancement of the quality-based clustering, proposed by Heyer (1999). The algorithm relies on a quality criteria which is defined in the beginning, and characterized by the minimum $r_{\min}$ and maximum distance $r_{\max}$ of a gene from its cluster center and the minimum number of members $n_{\min}$ for a cluster. A random center is formed within radius $r_{\max}$, that contain $n_{\min}$ members without adding data points that exceed r_{min}. If one data point exceeds $r_{\max}$, a new cluster is formed until all data points are assigned to a cluster with regards to $n_{\min}$ and $r_{\max}$. (Theresa Scharl et al., 2006). By the random initialization a non-deterministic approach is formed. Herein lies the difference to the originally proposed algorithm by Heyer (1999), where an original data point was used as start center. Nevertheless, no different local optima can be found and the algorithm always converges into the same optimum (Theresa Scharl et al., 2006). By the definition of r_{max} a certain cluster quality is ensured and profiles that do not fit are discarded. A measure for the similarity of two expression profiles is e.g. the distance of two objects (genes). Within the variety of distance metrics, the Euclidean one is used most frequently (Drăghici, 2012). It describes the shortest distance between two points (genes) x_1 and x_2 as a straight connection (Mimmack et al., 2001) in an extended form of Pythagoras' equation:

$$D_{x_1,x_2} = \sqrt{\sum_{i=1}^{n}(x_{1,i} - x_{2,i})^2}. \tag{2.2}$$

where the sum is formed over all available data points n in time for gene x_1 and x_2. One disadvantage of absolute distance measure lies in the resulting large distance of two profiles which show the same relative behavior, but at different absolute niveaus. However,

if the values were related to a standard, same relative changes result in same metric distances. If using an euclidean distance correlation the data should follow a normal distribution. This means coming from RNA-Seq, any data derived from a transformed normalized count metric is required, such as rlog as described in chapter 2.4.5.

To improve the overall clustering results (based on the cluster evaluation results described in the next section and summed up in table A.1), the K-means algorithm will be run afterwards, initiated with the resulting clusters and corresponding centers from SQBC. The K-means algorithm is sensitive to its center initialization and if the initial random selection of the pure K-means algorithm is poor, it might lead to a local optimum that is not as good as the global optima (Laszlo et al., 2007).

If only interested in overall expression profiles (strength and direction of linear relationship), Pearson correlation would serve this goal best. The Pearson coefficient (ranging from -1 to 1) only takes into account the shape of a profile, but not the absolute values (Eisen et al., 1998). Thus, a clustering algorithm using this distance measure is invariant with respect to shift and scaling of the time series. By preserving the absolute level of differences, the Euclidean distance is the measure that obtains the best, balanced general solutions Iglesias et al. (2013).For these reasons the euclidean distance measure in conjunction with SQBC and K-means clustering is used in this work.

Cluster Evaluation

After the choice and use of an appropriate algorithm, the evaluation of the resulting clusters is regarded as one of the most important steps (Halkidi et al., 2001). Algorithms categorically find clusters, independent whether these clusters are really existing or the data is just randomly composed (Jain and Dubes, 1988). In general, the target of cluster analysis is to find clusters, whose members are close to each other, exhibit a certain compactness (Halkidi et al., 2001) and have a certain extend of separation among themselves (D'haeseleer, 2005). To evaluate the cluster compactness, the intra-cluster variance is tested, as proposed by Handl et al. (2005). For characterization of separation, the mean silhouette coefficient S (Rousseeuw, 1987) is calculated. The silhouette coefficient of a data point i is a ratio of (i) the difference of the data points average distance to all data points of the other clusters $D_{i,k}$ and the mean distance between the data point and the members of its own cluster $\overline{D}_{i,j}$ and (ii) the maximum of the two means as given by

$$S_i = \frac{\overline{D}_{i,k} - \overline{D}_{i,j}}{max(\overline{D}_{i,k}; \overline{D}_{i,j})}. \tag{2.3}$$

The value of S ranges from -1 and 1, where a positive value close to one indicates high similarity to its assigned cluster compared to adjacent clusters and a negative value marks outliers. If a high number of points have a low or negative value, then cluster configuration may have too many or too few clusters (Gupta et al., 2016). A further method for evaluating the cluster quality, is testing its ability of prediction. Therefore, Yeung et al. (2001) proposed the calculation of the figure of merit. The idea is based on the assumption, that biologically related genes, should exhibit similar expression patterns (Jiang et al., 2004). Therefore, the validation measure requires correlated data, which is provided by gene expression profiles (Handl et al., 2005). The method proposed by Yeung et al. (2001) removes one data point of the data set, clusters genes on the remaining data and measures the within cluster similarity of expression values of the left out data point. The root-mean square deviation of a profile at the left-out time point of each cluster member is calculated relatively to the mean of its cluster. A small value indicates a high predictive power and reliability of the chosen cluster method (Jiang et al., 2004).

Storage of Gene Expression Analysis Results

Even with scale-down systems on the rise, the majority of the fundamental knowledge about regulatory effects caused by external stimuli has been acquired from shaking flask experiments (Murray, Schneider, et al., 2003; Shimada et al., 2011; Traxler, Zacharia, et al., 2011). Based on such and other experimental datasets, different databases have been established to summarize the knowledge about transcriptional regulation of different microorganisms. For *E. coli* two databases are of major interest, RegulonDB (Salgado et al., 2004) and EcoCyc database (Keseler et al., 2013). These databases integrate biological knowledge of the mechanisms that regulate the transcription initiation, as well as the organization of the genes and regulatory signals into operons in the chromosome. The project of RegulonDB started 1998 with around 500 regulation mechanisms and due to continuous gathering of information, version 9.3 with over 3000 experimentally confirmed regulatory interactions among around 1200 genes have been curated and released in 2017 (Gama-Castro et al., 2016), which can be used for performance assessment. To transfer and expand the knowledge gained by shaking flask experiments, scale-up studies have been used to investigate cellular performance under large-scale conditions which unraveled impaired process performance. Application examples are given by transcript time series and monitoring of metabolic changes reflecting the stimuli of glucose (Löffler, Simen, Jäger, et al., 2016b; Löffler, Simen, Müller, et al., 2017; Simen et al., 2017a), nitrogen (Brown et al., 2014), oxygen (Wulffen, Ulmer, et al., 2017; Liu et al., 2017) or temperature

stress (Caspeta et al., 2009) of *E. coli*. Data like this, derived from transcriptome analysis as it is described in figure 2.2, are the basis of proper validated mathematical models. Transcript analysis even enabled the engineering of robust *E. coli* strains (Michalowski et al., 2017) by attenuating the level of the alarmone ppGpp, the inducer of the stringent response regulation program. The new host is able to maintain high glucose uptake rates even under non or slow growing conditions. However, signal transduction is highly networked in the cells which may cause the cross-interference of multiple stimuli. The coincidence of multiple stimuli in large-scale fermentation is the rule rather than the exception (Xu, Jahic, et al., 1999; Egli, 1991). Accordingly, multiple stimulus/response studies are likely to gain importance in the future.

2.5 Modeling Microbial Growth with Different Granularity

Based on proper analyzed experimental data sets, mathematical models can be derived to simulate the microbial behavior under different conditions with a varying level of detail. Following the well-known classification of Bailey (1998) microbial models can be divided into non-structured/structured and non-segregated/ segregated approaches. Non-structured/ Non-segregated approaches represent the simplest growth models assuming average cells without sub-cellular detail. Such models are typically applied for bioprocess design. For the sake of simplicity, they are also applied in agent-based modeling for tracking individual cells. The consideration of subpopulations or individual cell properties leads to segregated approaches which, thanks to the improving availability of experimental data, is gaining more and more attraction. Structured, non-segregated models are commonly used for implementing the sub-cellular details of metabolic and transcriptional regulation, compartmentation or signal transduction (Nielsen et al., 1991; Tang, Deshmukh, et al., 2017). They are computationally intensive but represent a powerful tool for predicting detailed cellular responses to extracellular stimuli. A simplified version of this classification was used as general modeling approach in this thesis. The most accurate approach are structured/segregated models, which for example describe the whole glycolysis process with reactions for each enzyme, depending on enzyme affinities and turn over rates (Chassagnole et al., 2002). These parameters are more difficult to identify but transferable to other conditions. However, models like this are limited in scale, due to the complexity of the cellular mechanisms and the single cell consideration, which results in a quadratic scaling problem. Structured, non-segregated models typically consist

of a rigid network structure and a set of rate expressions including sensitive parameters. Knowledge of the network structure, the kinetic equations and the parameters is key to identifying a proper model. Often, such structures are determined following the bottom-up approach, i.e. the statistically profound identification of correlations between the structuring elements based on experimental data. The bottom-up concept can be applied to merge already existing small-scale models into large models (Klipp, Nordlander, et al., 2005; Guido et al., 2006; Brandon et al., 2015). Alternatively, top-down approaches aim for the identification of model parameters for a given structure. Accordingly, the top-down approach is a powerful tool for deciphering details of pathway interaction with the network, provided that the given structure is correct (Chou et al., 2009; Erickson et al., 2017). Since only limited computational power is available a structured, non-segregated structure is used to describe the cell population in a agent-based modeling approach. The cellular regulatory network based on experimental data can be derived with different methods presented in the following section.

2.6 Dynamic Regulatory Networks

A regulatory network model is the collection of macromolecules and their interactions, which together control the level of gene expression in a given genome. Numerous cellular processes throughout different layers (metabolome, proteome, transcriptome and genome) are affected by regulatory networks (Machado et al., 2011). Because of this enormous complexity and multilayer communication it is a major challenge to adequately display the regulatory system of microorganisms with computational tools. In the following different methods, which challenge gene regulatory approaches are introduced shortly with focus on a standard deterministic approach, ordinary differential equation (ODE) models, that follow the law of mass action.

Dynamic Network Set-ups

The main approaches which roughly embrace the computational possibilities are listed below and can be divided into boolean approaches (Thomas, 1973; Kauffman et al., 2003; Davidich et al., 2008), probabilistic models (Qian et al., 2002; Turner et al., 2004; Chandrasekaran et al., 2010) or ODE models (Chassagnole et al., 2002; Bolouri et al., 2002; Kremling, Bettenbrock, et al., 2007; Lemuth et al., 2008; Hardiman, Lemuth, Siemann-Herzberg, et al., 2009). An overview of the mentioned approaches is provided in figure 2.3. Furthermore different combinations of these methods have been applied to improve

the accuracy of the dynamic regulatory modeling (Shmulevich et al., 2002; Schlitt et al., 2007; Li, Li, et al., 2011).

Boolean Models

Regulatory networks are basically boolean networks with *on/off* conditions for all participating transcription factors and genes. Orth et al. (2010) stated, that boolean regulatory networks, besides ODE models, are the most used method to build regulatory regimes. If a gene is switched *off* in the regulatory network its respective flux is constrained to zero. If the gene is *on* the associated flux is unconstrained. Boolean networks were used to analyze the relationship between regulation functions and network stability in the yeast transcriptional network, using only the network's structure (Kauffman et al., 2003). It showed that boolean models do not correctly model the dynamics of a transcription factor that down-regulates its own expression, due to the model's limited level of detail (figure 2.3). In order to capture dynamic changes in the flux distribution of the metabolic network ordinary differential equations (ODE) are applied for different synthesis rates like mRNA formation or protein synthesis (Orth et al., 2010; Wulffen, Buchholz, et al., 2015). Another problem is the highly complex interplay between transcription factors, therefore it is hard to identify and model the contribution of individual factors satisfactorily (Wulffen, RecogNice-Team, et al., 2016). Mochizuki (2005) showed that the predictions of Boolean models can become unrealistic or too complex for larger networks when compared to those of the corresponding ODE models. However, when the number of entities is small and only qualitative knowledge is available, Boolean networks can provide important insights, such as the existence and nature of steady states or network robustness (Karlebach et al., 2008).

Probabilistic Models

The stochastic approach uses the inherent random nature of microscopic molecular collisions to build a probabilistic model of the reaction kinetics (figure 2.3). In particular, Gillespie's stochastic simulation algorithm (SSA) (Gillespie, 1976; Gillespie, 1977; Berry, 2002) is used to derive the behavior of single molecules. SSA takes as input the initial number of molecules of several species (for example, mRNAs and proteins) and reaction-probability constants and simulates the dynamics of the system, one reaction after another. A reaction probability is the probability that the necessary combination of specific molecules will participate in that reaction in an infinitesimal time interval (Karlebach et al., 2008).

The basic assumption of the algorithm is that the system is 'well stirred' meaning that each molecule always has an equal chance of being anywhere in the system's volume. This method was sufficiently described in Wilkinson (2011) and was successfully applied of Nieß et al. (2017) to describe the transcription of the *trp* operon during repetitive nitrogen starvation. Besides the high level of detail the drawback of this method is the large amount of data needed to describe every single molecule in the system and consequently the high computational effort.

Ordinary Differential Equation Models

A more general but also detailed model of regulation can be described by ODEs, as long as only one variable (e.g. space or time) change is considered (for changes in time and space, partial differential equations are appropriate). These equations continuously describe the change of entities over time (figure 2.3). In the context of transcriptomic networks, they describe the rate of change of transcript concentrations. The dynamic models based on an ODE system contain rate law equations for the reactions as well as their kinetic parameters and initial transcript concentrations. Building this type of model requires insight into time series dynamics to select appropriate rate laws, as well as experimental data for parameter estimation (Gennemark et al., 2009; Wang, Wang, Zhang, et al., 2007). Therefore, their application is more limited. Klipp, Herwig, et al. (2008) provides a good overview for the use of ODEs in biological context and gives some examples. Basic models of dynamic networks of well-studied organisms such as *E. coli* (Chassagnole et al., 2002, etc.) and *Saccharomyces cerevisiae* (*S. cerevisiae*) (Rizzi et al., 1997; Chen, Calzone, et al., 2004) are already publicly provided. The dynamic formulation needs significant information in terms of rate constants and total transcript concentrations, as well as reaction mechanisms to give rate laws and initial concentration levels, but generally rewards that effort with unique and detailed solutions. Thus, ODE models can generate predictions that may subsequently be compared to cellular phenotypes. Also Kremling, Heermann, et al. (2004) and Turner et al. (2004) stated that between stochastic modeling approaches for a low number of molecules (around 100) hardly any differences could be detected compared to the deterministic approach. Therefore high accuracy coupled to a comparably average computational effort is achieved.

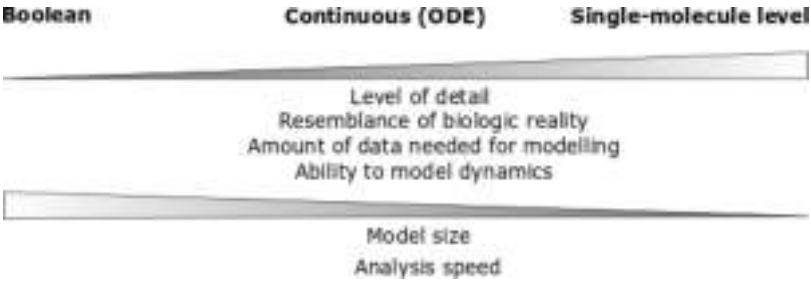


FIGURE 2.3: Summary of model approaches with relevant criteria. Models are listed along an imaginary scale, in which the amount of detail increases, from left to right. Several relevant criteria are indicated below the scale.

2.7 Computational Fluid Dynamic Simulations of Large - Scale Conditions

To predict cellular responses in large scale bioreactors, two crucial prerequisites are needed: (i) models simulating spatially resolved substrate availabilities, flows and mass transfer and (ii) models translating micro-environmental heterogeneities into proper cellular responses. As the second point was already introduced in the previous chapters, this chapter focuses on methods used in this thesis to simulate the environmental heterogeneities under large-scale conditions using a CFD approach.

Large-scale bioreactor conditions need to be calculated, aiming at a spatial resolution of mass, momentum and energy balances via numerical simulations. However, industrial processes of stirred tanks are usually operated in turbulent flow regimes (Re = $\frac{u\rho L}{\eta_l} > 10^4$) and under moderate conditions (30°C < T < 37°C; 1 bar < p < 1.5 bar) with media compositions ranging from low to highly viscous. Thus, several assumptions can be made to reduce the computational costs. In particular, for high Reynolds numbers (Re) friction is negligible and the general Navier-Stokes equation (NSE) resolves to the Euler equations. Additionally, the compressibility of a flow is directly related to the Mach number (Ma = $\frac{u}{c}$). When Ma < 0.3, the change in density in a given flow is minor and can be neglected. Hence, the flow can be considered incompressible ($\frac{D\rho}{Dt} = 0$). Incompressibility is also considered for the fluid simulated as the main components of media are water and salts. Additionally, the process in this study is considered isotherm and therefore the energy equation can be neglected. Hence, the Navier-Stokes and continuity equation can be written in Cartesian coordinates as:

$$\underbrace{\frac{\partial u_i}{\partial t}}_{\text{unsteady term}} + \underbrace{u_j \frac{\partial u_i}{\partial x_j}}_{\text{convective term}} = \underbrace{-\frac{1}{\rho}\frac{\partial p}{\partial x_i}}_{\text{pressure gradient term}} + \underbrace{\frac{\eta}{\rho}\frac{\partial^2 u_i}{\partial x_j \partial x_j}}_{\text{diffusive term}} + \underbrace{f_i}_{\text{external forcing term}} \tag{2.4}$$

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{2.5}$$

with u as fluid velocity, p as pressure, η as dynamic viscosity and ρ as density. Here, the external forcing term is considered as gravitational force (ρg_i) with g_i as gravitational acceleration.

Basically, NSEs describe the motion of viscous fluid flows with the fluids considered as a continuum rather than colliding particles. The formation of turbulent zones is provoked by eddies as a consequence of mixing. They are integrated via additional transport equations. Flows involving heat transfer (or compressibility) were not part of this thesis. Furthermore, the balancing of individual species (particles or cells) that undergo mixing and reactive changes require the implementation of proper conservation terms. Depending on the available computing power simulations can range from direct numerical simulation (DNS) via large eddy simulation (LES) to Reynolds-averaged Navier-Stokes (RANS) approaches. DNS yields to solve each individual turbulent structure, whereas LES only directly solves the large energy containing scales while the effects of the more universal small scales are modeled. RANS comprise time-averaged flow equations which allow to simulate small and large scale eddies with a minimum – but still challenging – computational effort (see figure 2.4; Ahlstedt et al., 1996; Zieringer and Takors, 2018). Consequently, RANS simulations are often favored. They only require 1/10 to 1/100 computational efforts compared to LES (Breuer et al., 1996) or DNS, respectively. Although RANS models require several modeling assumptions and approximations, their predictive power is sufficient for providing insight in reactor-scale substrate concentration gradients (Larsson, Törnkvist, et al., 1996), which is why a RANS approach was used in this study.

The RANS equations are time-averaged equations of motion for turbulent flows approximating different turbulent scales through fluctuating quantities, an idea first proposed by Reynolds (Reynolds, 1895). RANS models offer the most economic approach for simulating complex turbulent flows, because turbulences are considered with different levels of complexity. The most common RANS turbulence models are classified with respect to the number of additional transport equations that need to be solved along with

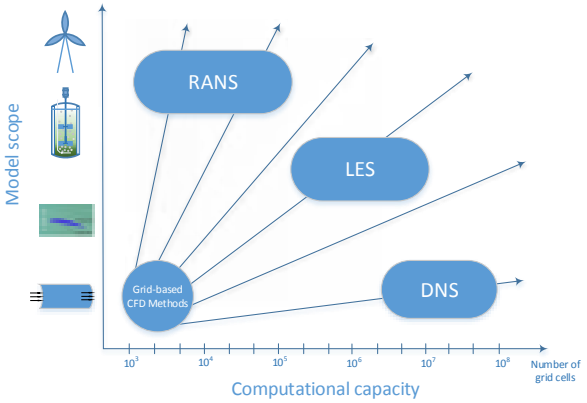


FIGURE 2.4: Different approaches of turbulence models, regarding the model scope and fields of application, as well as the corresponding computational capacity required. RANS: Reynolds-averaged Navier-Stokes equation, LES: Large Eddy Simulation, DNS: Direct Numerical Simulation.

the RANS flow equation. Besides, the often used two-equation models, such as the standard k-ϵ, k-ω or Renormalization group (RNG) k-ϵ models, one-equation models (low-cost RANS models, e.g. the Spalart-Allmaras approach) or even zero-equation models which estimate the turbulence viscosity via the mean velocity and the length scale using an empirical formula are available (Rodi, 2017).

2.7.1 Reynolds-averaged Navier-Stokes Two Equation Turbulence Models

RANS models are usually employed to simulate turbulent industrial flows (Corson et al., 2009). This method averages the equations of motion over time to reduce the computational effort and complexity of the model. Using the approximation of the RANS approach the NSE gives, as momentum equation for incompressible flows:

$$\frac{\partial(\rho\overline{u}_i)}{\partial t} + \frac{\partial}{\partial x_j}(\rho\overline{u}_i\overline{u}_j + \rho\overline{u_i'u_j'}) = \frac{\partial\overline{p}}{\partial x_i} + \frac{\partial}{\partial x_j}\left[\eta\left(\frac{\partial\overline{u}_i}{\partial x_j} + \frac{\partial\overline{u}_j}{\partial x_i}\right)\right] \tag{2.6}$$

where u is the velocity of the fluid with index i as directions in space, ρ is the fluid density, η is the fluid viscosity, p is the pressure and the turbulent fluctuations ${u_i}'$. When it comes to turbulence involved in the fluid flow additional non linear terms (Reynolds

stress terms) appear, which can be computed in varying degrees of complexity. Determining the right choice of a turbulence model depends on the detail of results expected and computational capability available. The most common RANS turbulence models are classified on the basis of the number of additional transport equations that need to be solved along with the RANS flow equation. The Reynolds stress terms can be closed via the Boussinesq hypothesis as first order closure model, where the Reynolds stresses are modeled using an eddy viscosity η_t or via transport equations for Reynolds stresses as second order closure model (RSM). The RSM is more complex, computationally intensive and more difficult to converge than linear eddy viscosity models. Therefore only the k-ϵ two equation model was used to simulate turbulence in a industry scale stirred tank reactor.

Standard and Realizable k-ϵ Model

The k-ϵ model is the most widely used engineering turbulence model for industrial applications, because of its robustness and accuracy (Fluent, 2009). The standard k-ϵ model is essentially a high Reynolds number model and assumes the existence of isotropic turbulence and spectral equilibrium. The model focuses on transport and dissipation of kinetic energy with two additional transport equations (k and ϵ). The transport equations for turbulent kinetic energy k and dissipation rate ϵ are solved so that turbulent viscosity can be computed for RANS equations. First to be introduced is the turbulent kinetic energy which is calculated as kinetic energy per unit mass of the turbulent fluctuations $u_i{}'$:

$$k = \frac{1}{2}\overline{u_i' u_i'} = \frac{1}{2}\left(\overline{u_x'^2} + \overline{u_y'^2} + \overline{u_z'^2}\right) \tag{2.7}$$

The dissipation rate ϵ is the rate at which the turbulent kinetic energy k dissipates and is converted into thermal internal energy. It can be described physically as:

$$\epsilon = 2\nu\overline{e_{ij}' e_{ij}'} \qquad \text{with} \qquad \overline{e_{ij}'} = \frac{\partial u_i'}{\partial x_j} + \frac{\partial u_j'}{\partial x_i} \tag{2.8}$$

The transport equation for k is defined as follows, without the generation terms caused by buoyancy and compressible effects:

$$\frac{\partial(\rho k)}{\partial t} + \nabla(\rho k u) = \nabla\left[(\eta + \frac{\eta_t}{\sigma_k})\nabla k\right] + G_k - \rho\epsilon \tag{2.9}$$

where the generation of turbulent kinetic energy due to mean velocity gradients is indicated as G_k, the turbulent Prandtl number as σ_k and the Nabla operator as ∇ (derivative in all spacial directions). G_k can be expressed with the Boussinesq approach:

$$G_k = -\rho\overline{u_i' u_j'}\frac{\partial u_i}{\partial x_j} \qquad \text{where} \qquad -\rho\overline{u_i' u_j'} = -\rho\frac{2}{3}k\partial_{ij} + \eta_t\left(\frac{\partial \overline{u_i}}{\partial x_j} + \frac{\partial \overline{u_j}}{\partial x_i}\right) \tag{2.10}$$

The Boussinesq hypothesis states, that the Reynolds stresses could be linked to the mean rate of deformation (like it is done with viscous stresses), where δ_{ij} is the Kronecker delta, defined as:

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} \tag{2.11}$$

The transport equation for ϵ otherwise is described as

$$\frac{\partial(\rho\epsilon)}{\partial t} + \nabla(\rho\epsilon u) = \nabla\left[(\eta + \frac{\eta_t}{\sigma_\epsilon})\nabla\epsilon\right] + C_{1\epsilon}\frac{\epsilon}{k}G_k - C_{2\epsilon}\rho\frac{\epsilon^2}{k} \tag{2.12}$$

with the empirical constants $C_{1\epsilon}$, $C_{2\epsilon}$ (see table 2.1), the turbulent Schmidt number σ_ϵ and η_t as turbulent viscosity (eddy viscosity), which is determined from a single turbulence length scale

$$\eta_t = \rho C_\eta \frac{k^2}{\epsilon} \tag{2.13}$$

where C_η is another empirical constant (table 2.1). This artificial term is isotropic and describes the phenomenon that a highly turbulent system behaves similarly to a more viscous system. Isotropic in this case means that the ratio between Reynolds stress and the mean rate of deformation is the same in all directions. This assumption fails in many categories of flow where it leads to inaccurate flow predictions and can only be overcome by deriving and solving transport equations for the Reynold stresses themselves (as it is applied in the computationally more expensive RSM) (Versteeg et al., 2007).

TABLE 2.1: Empirical Constants based on Launder et al. (1972)

$C_{1\epsilon}$	$C_{2\epsilon}$	C_η	C_k	C_ϵ
1.44	1.92	0.09	1.0	1.3

The realizable k-ϵ models differs from the standard model in formulation of the turbulent viscosity η_t and the transport equation for ϵ derived from an exact equation for the

transport of the mean square vorticity variations (Shih et al., 1994), which reads as:

$$\frac{\partial(\rho\epsilon)}{\partial t} + \nabla(\rho\epsilon u) = \nabla\left[(\eta + \frac{\eta_t}{\sigma_\epsilon})\nabla\epsilon\right] + \rho C_1 S\epsilon - \rho C_2 \rho \frac{\epsilon^2}{k_t + \sqrt{v\epsilon}} \tag{2.14}$$

with $C_1 = \max(0.43, \frac{\eta}{5+\eta})$, $\eta = \frac{Sk_t}{\epsilon}$ and $S = \sqrt{2S_{ij}S_{ij}}$. With an improved equation for ϵ and therefore improved performance for flows with rotation, recirculation, strong streamline curvature, boundary layers under strong adverse pressure gradients or separation, the realizable k-ϵ model performs better, which makes it very well applicable for the studied system (Shih et al., 1994).

The standard as well as the realizable k-ϵ model seems to be best for free shear layer flows with small pressure gradients. Especially when the model is combined with additional wall functions, the greatest weakness of this model can be overcome (Patel et al., 1985). It is also the most popular two-equation turbulence model in use today. When utilized in conjunction with wall functions or multiple reference frame (MRF) for impeller rotation (Coroneo et al., 2011), the k-ϵ model is reasonably well behaved and has been applied to the solution of a variety of engineering problems with a good amount of success (Lapin, Klann, et al., 2010; Wutz et al., 2016; Haringa, Deshmukh, et al., 2017; Kuschel, Siebler, et al., 2017).

2.7.2 Impeller turbulence

Turbulence in stirred tank reactors is mainly induced via the motion of impellers. Several different methods exist to impose motion where the impeller is fixed (e.g. MRF (Luo and Gosman, 1994)) or varies (e.g. Sliding mesh (SM) (Murphy, 1994)). Fixed impeller position, where motion is imposed either at the impeller or in the surrounding zones yields a steady state solution valid for only this explicit position whereas the SM method models a transient behavior of the fluid motion, but inquires higher computational costs (Dewan et al., 2006). Although the SM method captures complex 3D flow structures, the MRF method is used most frequently due to its lower computational burden and good enough agreement with experimental data (Naude et al., 1998; Tabib et al., 2017; Sommerfeld et al., 2004). Haringa, Vandewijer, et al. (2018a) experimentally revealed the presence of macro-instabilities in the region between the impellers, as well as a peak in the turbulent kinetic energy in the region where the flow from the individual impellers converges. The MRF-RANS method was found unable to capture both. To improve the mixing results a SM-RANS approach was used which captures the effect of macro-instabilities, while only mildly underestimating the turbulent kinetic energy k_t (Haringa, Vandewijer,

et al., 2018b). The results show good accordance to experimental data of Laser-Doppler Anemometry (LDA) (Ng et al., 1998). To counteract this deficiency in k_t via η_t, the turbulent Schmidt number can be tuned ($Sc_t = \frac{\eta_t}{\rho D_t}$ = 0.1 - 0.2) (Montante et al., 2005; Gunyol et al., 2009). Hence, the computational mesh was divided into stationary and rotating blocks connected to each other with a sliding surface.

2.8 Modeling of Euler- Euler and Euler- Lagrange Flows in Stirred Tank Reactors

To investigate the consequences and appearance of environmental heterogeneities, proper modeling frameworks should link local variations with cellular kinetics and sufficiently resolve droplet size distributions. Regarding multiphase modeling, three common approaches exist: Volume-of-fluid (VOF), mixture and Eulerian model. VOF simulates phase boundary explicitly, that means when expecting a lot of small droplets (as it is true for oil) an extremely fine mesh is necessary for sufficient resolution of the droplet surface. VOF is therefore mostly applied for free-surface flows. Easier to simulate by only using an averaged content of both phases in each grid element (quasi-continuous) is the mixture model. It is less computationally costly and applicable if the dispersed phase is widely distributed in the fluid phase. Regarding oil distribution in a stirred tank reactor, droplets are often concentrated at the top of the liquid surface or around the impellers. Additionally, interphase forces play an important role which makes the Eulerian model the tool of choice for bi-phasic simulation (oil-in-water). Based on the grid-based Euler approach mainly two different methods exist to display flow characteristics: (i) population balance models (PBM) and (ii) Euler-Lagrange method (EL).

Euler-Euler Approach

Regarding the Euler-Euler approach, both phases are considered as a continuum and two separate momentum equations are solved, with closure relations resembling interphase momentum transfer. To be consistent with the governing equations of Navier-Stokes, an averaged form of the Eulerian representation was derived. Applying phase averaging to alleviate the complication of modeling individual phase entities, the following equation for phase k can be formulated (Yeoh et al., 2014):

$$\frac{\partial \alpha_k}{\partial t} + \frac{\partial (\alpha_k \overline{u_i}^k)}{\partial x_i} = \Gamma_k \tag{2.15a}$$

$$\rho_k \left(\frac{\partial (\alpha_k \overline{u_i}^k)}{\partial t} + \rho \frac{\partial (\alpha_k \overline{u_i}^k \overline{u_i}^k)}{\partial x_j} \right) = -\alpha_k \frac{\partial \overline{p}}{\partial x_i} + \frac{\partial (\alpha_k \overline{\tau_{ji}}^k + \overline{\tau_{ji}}^{Re,k})}{\partial x_j} - \alpha_k \rho_k g_i + M_{i,k} \tag{2.15b}$$

$$\tau_{ji}^{Re,k} = -\rho_k \overline{u_i'^k u_j'^k} \tag{2.15c}$$

with α_k as volume fraction ($\alpha_L + \alpha_g = 1$), Γ_k as source/sink term to describe mass transfer, $\tau_{ji}^{Re,k}$ as Reynolds stress term and $M_{i,k}$ as interphase momentum exchange (drag, lift, virtual mass, turbulent and wall lubrication forces). If a size distribution is simulated, typically only the Sauter diameter d_{32} is considered for interfacial momentum exchange:

$$d_{32} = \frac{\sum_{i=1} n_i d_i^3}{\sum_{i=1} n_i d_i^2} \tag{2.16}$$

meaning a monodisperse system with a droplet diameter of d_{32} has the same volume-related interfacial area as a particle size distribution of the disperse phase in a real system with d_i as particle diameter and n_i as number.

A continuum is a continuous system in which adjacent elements do not perceptibly differ from each other although the endpoints of the system may be drastically distinct (Schmalzriedt et al., 2003). However, microorganisms are individual in their behavior, and therefore the description as continuum is not completely correct. In fact, the continuum approach leads to loss of the individual responses of the cells, for example when considering starvation effects during fed-batch fermentations. The same accounts for the dynamic change of droplet size distributions, where a constant diameter would only poorly describe the droplet behavior in the reactor. In some instances, these can be tackled by combining the Euler approach for the fluid phase with Population Balance Equations (PBE) (Bezzo et al., 2003; Morchain et al., 2014; Heins et al., 2015; Wang, Wang, and Jin, 2005). Using this approach, particles (cells or droplets) are grouped in classes and a predefined distribution range of particles is implemented via distribution density functions including growth, coalescence and break-up terms. Thus, a complete particle size distribution can be represented by a limited number of parameters, which drastically limits the computation effort. These equations are useful to determine relevant macroscopic properties such as the interfacial area or the biomass-specific growth rate (Sajjadi et al., 2013; Venneker et al., 2002; Haringa, Deshmukh, et al., 2017). For a multiphase system

of nonreactive, isothermal, particle mixtures, it is customary to assume that all relevant internal variables can be calculated from the consideration of the particle volume or diameter (Yeoh et al., 2014). A simplified overview of the intersection of the two fluid model and PBM is provided in figure 2.5.

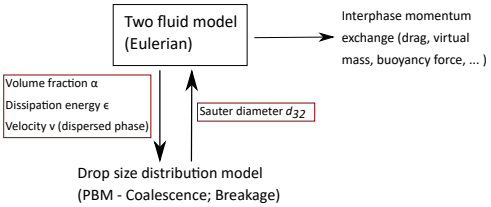


FIGURE 2.5: Interaction of continuous two-fluid Eulerian model with PBM model.

Considering the special case, where there is no mass change of the particle (absence of particle growth), the transport equation for the particle number density reduces to

$$\frac{\partial}{\partial t}[n(V,t)] + \frac{\partial}{\partial x_i}[u_i n(V,t)] = \frac{1}{2}\int_0^V a(V-V',V')n(V-V',t)n(V',t)dV' - \int_0^{\inf} a(V,V')n(V,t)n(V',t)dV' + \int_{\omega_v} g(V')\beta(V|V')n(V',t)dV' - g(V)n(V,t) \tag{2.17}$$

Birth due to Coalescence; Death due to Coalescence; Birth due to Breakage; Death due to Breakage

PBMs represent a powerful modeling framework for the description of fundamental properties that are characterized by distributions in a coarse time scale. However, cellular adaptations may happen on different timescales than macroscopic fluctuations and may show much more interactions than implemented in common PBMs. Another inherent limitation of the PBE approach is that the incorporation of a detailed intracellular reaction network leads to massive computational effort because a high dimensional distribution function must be computed. Additionally, no information on the level of single particles can be provided within this approach. To overcome those limitations the Lagrangian method is applied to receive single particle trajectories (Lapin, Müller, et al., 2004).

Euler-Lagrange Approach

Besides the Euler-Euler approach, there is a second widely used method which treats only the liquid phase motions in an Eulerian representation and computes the motion of the dispersed phase fluid elements in a Lagrangian way by individually tracking them on their way through the reactor. Individual properties are thereby assigned to each moving particle (e.g. biological cell). Noteworthy, these traits cause interactions with the surrounding broth receiving equal environmental feedback. In essence, EL tracks a given number of particles while fluctuating in the bioreactor thereby recording interactions with the environment. The dispersed phase can exchange momentum, mass, and energy with the fluid phase. The approach renders considerably simpler when particle-particle interactions may be neglected. This requires a fully diluted (volume fraction $< 10\,\%$) dispersed second phase. Considering the case of biological cells as moving particles, further simplifications are often assumed: Since the observation window of fluctuating cells is smaller than time constants of physical changes inside the grid, cellular impacts on physical states are often neglected. Additionally, individually simulated particles account for cell ensembles with identical trajectories as introduced by Lapin, Müller, et al. (2004).

Each particle has its own coordinates and properties whereas the fluid phase is further regarded as continuum. The Eulerian phase is being solved through the Navier-Stokes equation and the disperse phase is solved through tracking of the single bubbles or particles. To describe the transport processes across the particle membrane (in this case a microorganism cell membrane), intracellular balance equations are required, i.e. substrate uptake rate and product excretion. The numerical solution of the resulting system is limited by the immense computing power required for three-dimensional discretization in space and the necessity to populate each of the resulting finite volumes with a sufficient number of cells to minimize the effect of statistical error on the accuracy of the solution (Lapin, Müller, et al., 2004; Haringa, Tang, Deshmukh, et al., 2016). That is why particles are often regarded as observers rather than creator of their environment. The particle history can be recorded and exported serving as a input for further analysis in a different framework.

2.9 Oil in Bioprocesses

In microbial cultivation processes, productivity advances are frequently limited by the transport of substrate or (by-) product into or out of the reactor. Oxygen is one example of a sparingly soluble substrate whose transport is a critical factor in aerobic cultures.

Due to this crucial impact on productivity, many different solutions have been proposed to increase oxygen transfer rates. One way to alleviate this transport limitation is to add a second, water-immiscible phase in which oxygen has a greater solubility and can act as additional oxygen carriers releasing oxygen depending on its surrounding concentration. For example, sunflower oil with a 5-times higher solubility of oxygen than water (water: 8 ppm) (Cuvelier et al., 2017; Zlokarnik, 2013). Additionally, some bioprocesses use additives to facilitate down-stream analysis like product purification or to prevent cell-product interactions like product degradation or inhibition. In this context, oils (organic phase) often serve as hydrophobic product carrier (Wubbolts et al., 1996; Becker, Puel, et al., 2014; Schmid et al., 1998; Patil et al., 2020). According to McMillan et al. (1987) a 250 % increase in the amount of product produced per fermentation in a oxygen-limited process at a oil volume fraction of 0.4 might be achievable, which shows the significant potential for increasing productivities using oil-in-water dispersions.

This side-project specifically focuses on oxygen mass transfer in a oil in water system and the oil droplet distribution in a stirred tank reactor. Opposite to the examples above, not all oils have a positive effect on mass transfer and often reported $k_L a$ values are contradictionary, preventing any definite conclusion being reached (Dumont et al., 2003; Kaur et al., 2007). There have been three distinct approaches reported in literature to explain the change in mass transfer in a gas–liquid–liquid system: interfacial blockage, shuttle mechanism and permeability effect (Dumont et al., 2003).

Interfacial blockage: Interfacial properties (e.g. variation in the interfacial area and the mass transfer coefficient) of the oil-in–water system can be expressed through the value of the spreading coefficient (S) determining whether a (organic) liquid droplet will spread initially on the gas interface (S > 0) or not (S < 0) (Yoshida et al., 1970). Liquids with a positive spreading coefficient (S > 0) enhance the gas absorption rate, whereas liquids having a negative spreading coefficient (S < 0) retard the gas absorption. The spreading coefficient is relatively easy to calculate by only relying on surface tensions. S is defined as:

$$S = \sigma_{w/g} - (\sigma_{o/g} + \sigma_{o/w}) \tag{2.18}$$

However, it is not possible to explain the oil distribution near the gas–liquid interface through this parameter only (Yoshida et al., 1970; Dumont et al., 2003).

Shuttle mechanism: Oil droplets absorb oxygen in the gas-aqueous boundary layer and transfer the oxygen to the aqueous phase when they have returned to the bulk liquid. The major assumption of this mechanism is that oil droplets are smaller than the film thickness (5-10 µm). The presence of very fine, gas-absorbing droplets within the mass transfer zone near the gas-liquid interfaces may increase the specific gas absorption rate (Brilman, 2000). There is no direct evidence to prove the shuttle mechanism: the theory is based on experimental evidence that shows that the gas transfer rate is dependent on oil droplet size. For more information on this mechanisms, see Dumont et al. (2003), Kaur et al. (2007), and Van Ede et al. (1995).

Permeability effect: Dynamic interaction of the oil droplets with the concentration boundary layer causing increased turbulence or mixing in this layer. Hence oxygen first diffuses across the thin aqueous layer and then into either the oil dispersed phase or the continuous aqueous phase (Dumont et al., 2003). Thus, oxygen permeability through an organic layer can be significantly greater than through an aqueous layer of equivalent depth. According to Mcmillan et al. (1990), gas permeability, which takes into account both gas solubility and gas diffusivity, is a more accurate index of the capacity to transport solute than solubility alone. Indeed, in the case of high - viscosity oils, the outcome of high gas solubility can be inadequate owing to low gas diffusivity. As shown in table 2.2 safflower oil is a rather less viscous oil, thus diffusivity is not hindered and the remaining two effects might cover the mechanisms of mass transfer in oil in water systems sufficiently.

Additionally, the **partitioning coefficient** describes the ratio of concentrations of a compound in a mixture of two immiscible solvents at equilibrium. Therefore, this ratio is a comparison of the solubilities of the solute in these two liquids. Partitioning coefficient was not determined in this study, but might be of interest for further investigations. To determine oxygen mass transfer rate the dynamic method without organisms has been chosen, which is explained in Appendix E.

2.10 Safflower Oil

Studies of Patil et al. (2020) brought safflower oil as product carrier and $k_L a$ enhancer to the attention of this study. As only little is known about the effect of safflower oil on bubble size distribution and oxygen mass transfer, some basic studies are performed

as part of a side-project in this thesis. This chapter summarizes the main properties of safflower oil.

Cheap in production and sustainable thus gained from regenerative natural source it is a good alternative to often used silicon oils (energy intensive production from silica sand). It is a stable and high-quality oil consisting of long chain, polyunsaturated fatty acids. The main characteristics of safflower oil can be withdrawn from the following table:

TABLE 2.2: Physical properties of safflower oil at 25°C (Krist, 2013)

Property	Value	Unit
Density	0.922- 0.938	kg m^{-3}
Viscosity	0.0299 (at 38 °C)	Pas
Index of refraction	1.473- 1.475	
Saponification value	186- 203	mg_{KOH} $g_{oil}{}^{-}1$
Unsaponifiable share	0.5- 1.5	%
Melting point	-5	°C
Solidification point	-13 to -20	°C

Unfortunately, oxygen transport capacities (e.g. perfluorohydrocarbons (high affinity towards oxygen): 0.17 g_{oxygen} $(Lh)^{-1}$) and oxygen solubility in safflower oil (e.g. sunflower oil: 40 ppm) is still unknown. So far, safflower oil is mainly known from pharmaceutical to body care purposes, rather than biotechnological application which is only used recently (Krist, 2013; Patil et al., 2020). However, the method of extracting fat-soluble products with oils is already familiar (by mostly using silicon oils; Thirumangalathu et al., 2009).

Chapter 3

Author contributions

Any of the attached publications were prepared in co-authorship with other researchers as indicated in the following paragraphs. Smaller contributions are declared within each publication in the respective Acknowledgements section.

Research paper: Data-driven *in silico* prediction of regulation heterogeneity and ATP demands of *Escherichia coli* in large-scale bioreactors

Simulation, evaluation of results and writing of the manuscript was accomplished by Julia Zieringer. Julia Zieringer is first author of this manuscript. Moritz Wild implemented major parts of the agent-based model under supervision of Julia Zieringer. Prof. Dr.-Ing. Ralf Takors supervised the research and is the corresponding author.

Research paper: Transcriptional profiling of stringent response mutant strain *E. coli* SR reveals enhanced robustness to large-scale conditions

Writing of the manuscript was accomplished equally by Martin Ziegler and Julia Zieringer. Martin Ziegler and Julia Zieringer share co-first authorship. Prof. Dr.-Ing. Ralf Takors contributed to the manuscript's content by selected additions and reviewed it. Prof. Dr.-Ing. Ralf Takors is the corresponding author. Martin Ziegler planned and conducted all experiments in this study. Martin Ziegler collected and analyzed the primary experimental data. Julia Zieringer conducted the transcriptomic analysis. Prof. Dr.-Ing. Ralf Takors supervised the research and is the corresponding author.

Research paper: Engineering of robust *Escherichia coli* chassis strains for large-scale fermentations

Writing of the manuscript was accomplished by Martin Ziegler. Martin Ziegler is the first author. Prof. Dr.-Ing. Ralf Takors contributed to the manuscript's content by selected

additions and reviewed it. Prof. Dr.-Ing. Ralf Takors is the corresponding author. Martin Ziegler planned the genomic alterations in this study and partially conducted them. The remaining lab work was conducted by students supervised by Martin Ziegler. Martin Ziegler planned and conducted all shaking flask and fermentation experiments in this study. Martin Ziegler collected and analyzed the primary experimental data. Julia Zieringer prepared the genomic sequencing data and data analysis methods. Content analysis was conducted by Martin Ziegler. Prof. Dr.-Ing. Ralf Takors supervised the research and is the corresponding author.

Contributions to other data shown in this thesis

Data presented in this thesis which was not covered in any of the attached publications was analyzed by Julia Zieringer. Some research activities were conducted by students directly supervised by Julia Zieringer:

- Janina Faiß (Semester Thesis)
- Michaela Wick (Bachelor Thesis)
- Kimberly Fichtel (Bachelor Thesis)
- Moritz Wild (Master Thesis)

Contributions to other publications not shown in this thesis

For reasons of space, only the main publications have been included in this thesis. Further publications and contributions to other research papers are listed in the following:

- Zieringer, J., & Takors, R. (2018). *In silico* prediction of large-scale microbial production performance: constraints for getting proper data-driven models. Computational and structural biotechnology journal, 16, 246-256.
- Verhagen, N., Zieringer, J., & Takors, R. (2020). Methylthioadenosine (MTA) boosts cell-specific productivities of Chinese hamster ovary cultures: dosage effects on proliferation, cell cycle and gene expression. FEBS Open bio, 10(12), 2791-2804.
- Hajian, C. S. S., Zieringer, J., & Takors, R. (2020). Euler-Lagrangian Simulations: A Proper Tool for Predicting Cellular Performance in Industrial Scale Bioreactors.

Chapter 4

Engineering of robust *Escherichia coli* chassis and exploitation for large-scale production processes

Authors

Martin Ziegler (first author) - University of Stuttgart, Institute of Biochemical Engineering
Julia Zieringer (second author) - University of Stuttgart, Institute of Biochemical Engineering
Clarissa-Laura Döring (third author) - University of Stuttgart, Institute of Biochemical Engineering
Liv Paul (fourth author) - University of Stuttgart, Institute of Biochemical Engineering
Christoph Schaal (fifth author) - University of Stuttgart, Institute of Biochemical Engineering
Ralf Takors (corresponding author) - University of Stuttgart, Institute of Biochemical Engineering

4.1 Abstract

In large-scale bioprocesses microbes are exposed to heterogeneous substrate availability reducing the overall process performance. A series of deletion strains was constructed from *E. coli* MG1655 aiming for a robust phenotype in heterogeneous fermentations with transient starvation. Deletion targets were hand-picked based n a list of genes derived from previous large-sclae simulation runs. Each gene deletion was conducted on the premise of strict neutrality towards growth parameters in glucose minimal medium. The final strain of the series, named *E. coli* RM214, was cultivated continuously in an STR-PFR

(stirred tank reactor - plug flow reactor) scale-down reactor. The scale-down reactor system simulated repeated passages through a glucose starvation zone. When exposed to nutrient gradients, *E. coli* RM214 had a significantly lower maintenance coefficient than *E. coli* MG1655 ($\Delta m_s = 0.038\ _{Glucose}\ g_{CDW}^{-1}\ h^{-1}$, $p < 0.05$). In an exemplary protein production scenario *E. coli* RM214 remained significantly more productive than *E. coli* MG1655 reaching 44 % higher eGFP yield after 28 h of STR-PFR cultivation. This study developed *E. coli* RM214 as a robust chassis strain and demonstrated the feasibility of engineering microbial hosts for large-scale applications.

4.2 Introduction

Large-scale fed-batch bioprocesses often suffer from reduced process performance compared to lab-scale experiments conducted during process development (Bylund, Castan, et al., 2000; Enfors et al., 2001). The physical and engineering constraints in large-scale reactors inevitably lead to the formation of spatial heterogeneities in relevant process parameters such as nutrient availability, concentrations of dissolved oxygen, carbon dioxide and pH (Bylund, Collet, et al., 1998; Cortés et al., 2016). Heterogeneities of nutrient availability are caused by long mixing times of large-scale reactors (Delvigne, Destain, et al., 2006; Noorman, 2011). Studies employing computational fluid dynamics (CFD) have revealed that in fed-batch processes this typically results in the formation of zones with high nutrient concentrations close to the feeding point and zones depleted of nutrients at the far end of the reactor (Haringa, Deshmukh, et al., 2017; Kuschel and Takors, 2020). Depending on the mixing time and their position in the reactor cells frequently move through different zones on a timescale of seconds to minutes and cellular regulatory programs ranging from overflow metabolism to starvation responses are repeatedly triggered and shut down (Kuschel, Siebler, et al., 2017). Due to the delay of transcriptional responses, regulatory consequences of stress stimuli may be effective distant from the spot of stress induction which finally creates a heterogeneous population status (Nieß et al., 2017; Zieringer, Wild, et al., 2021). There is evidence from an increasing number of studies that the performance of many industrial workhorse organisms such as *Escherichia coli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Saccharomyces cerevisiae* and *Penicillium chrysogenum* is negatively affected when facing process heterogeneities (George et al., 1993; Junne et al., 2011; Larsson and Enfors, 1988; Vasilakou et al., 2020; Jonge et al., 2011; Olughu, Nienow, et al., 2020).

Substantial effort has been made by the scientific community to understand microbial responses to different zones occurring in large-scale reactors (Lara, Galindo, et al., 2006; Lara, Leal, et al., 2006; Olughu, Deepika, et al., 2019). In academic laboratories, the conditions of industrial reactors are commonly simulated using multi-compartment scale-down reactors (Delvigne, Takors, et al., 2017; Neubauer and Junne, 2010; Takors, 2012).Typically, nutrient pulsing or secondary vessels are employed to deliver a stimulus representative for the conditions under investigation (Bylund, Guillard, et al., 1999). The design of a scale-down reactor also serves to control the circulation of the microbial population and its residence time in stimulus zones. A commonly used design follows a two-compartment approach and consists of a primary stirred tank reactor (STR) coupled to a secondary plug-flow reactor (PFR). While the STR represents the bulk of the fermentation broth, the plug-flow compartment represents a stimulus zone with a defined residence time. Together, the STR-PFR two-compartment reactor enables the study of cellular behavior in heterogeneous environments.

Zones with low nutrient concentration but high oxygen availability occur in reactor segments far away from the feeding point. The effects of such transient starvation conditions on the performance and intracellular regulation of microbial populations can be studied in C limited scale down reactors. In the case of *Escherichia coli* K-12 repeated passages of cells through starvation zones were found to negatively impact process performance which could be observed as a reduced biomass yield (Neubauer, Häggström, et al., 1995). In parallel, regulatory responses such as the stringent response and the general stress response are rapidly initiated (Delvigne, Boxus, et al., 2009; Löffler, Simen, Jäger, et al., 2016b; Neubauer, Åhman, et al., 1995; Simen et al., 2017a; Sunya et al., 2012). Noteworthy, these cellular responses serve rather long term than short term needs and appear to be futile if cells enter zones of nutrient access shortly after the induction of the strategic precaution measure. Transcriptional investigations in a carbon limited STR PFR system offered a potential link between futile regulation and reduced process performance: Frequent transcriptional reprogramming was proposed to cause high secondary metabolic costs from aberrant transcription and translation (Löffler, Simen, Jäger, et al., 2016b). Earlier studies employing *Escherichia coli* K-12 strains in two-compartment scale-down reactors have indicated that the repeated passage of cells through a starvation zone impedes process performance (Neubauer, Häggström, et al., 1995). Refined experiments suggested high secondary metabolic costs as cellular regulation frequently initiates transcriptional reprogramming as the underlying cause (Löffler, Simen, Jäger, et al., 2016b). It

was estimated that an increased maintenance of up to 30 – 40% was caused by the transcriptional oscillations and a substantial fraction of this originated from the expression of open reading frames whose products appeared to bestow no apparent benefit in a controlled bioprocess employing standard glucose minimal medium.

The data collected by Löffler, Simen, Jäger, et al. (2016b) led us to propose a novel design approach for production strains: We reasoned that an intelligently engineered deletion strain might have advantages in conditions that repeatedly induce wasteful expression of process irrelevant genes. A heterogeneous fermentation with repeated transient starvation could then be a suitable testing environment. The choice of deletion targets would have to be based on the estimated effect of the deletion and be restricted by the requirements of neutrality towards growth and global regulation. The design process differs from previous considerations on the creation of lean proteome strains in the regard that savings only become apparent due to fluctuating induction (Valgepea et al., 2015). Secondary metabolic costs can traditionally be assessed through Pirt's maintenance coefficient (Pirt, 1965). We hypothesized that the deletion of a suitable set of genes should lead to a reduced maintenance coefficient under scale down conditions representing starvation zones. The resulting strain could then serve as a base strain for the construction of robust production strains.

We identified deletion candidates matching the defined criteria and constructed a series of deletion strains from *E. coli* MG1655. The final strain of the series, named *E. coli* RM214, was fermented in continuous cultivations in an STR-PFR system simulating starvation zones. *E. coli* RM214 had a significantly lower maintenance coefficient than *E. coli* MG1655 under simulated large-scale conditions. We then characterized *E. coli* RM214 in an exemplary protein production scenario using eGFP as a model product. Compared to *E. coli* MG1655, the deletion strain showed an increased resilience towards the scale down conditions as evidenced by reduced productivity losses and a higher fraction of producing cells.

4.3 Materials and Methods

4.3.1 Bacterial Strains, Media, and Buffer Solutions

All strains used in this study are listed in table 4.1.

2xYT medium was prepared by autoclaving 16 gL^{-1} tryptone, 10 gL^{-1} yeast extract, 5 gL^{-1} NaCl dissolved in demineralized water. For agar plates 18 gL^{-1} agar-agar was added prior to autoclavation. For ph indicator plates 0.03 gL^{-1} of neutral red and 10

gL^{-1} Rhamnose were supplemented from sterile stock solutions directly before pouring. SOC medium was prepared as described previously (Hanahan, 1983). Agar plates for *tetA-sacB* counterselection were prepared as described previously (Li, Thomason, et al., 2013). If strains with antibiotic resistance markers were cultivated, antibiotics were added to media in the following concentrations: Chloramphenicol 20 µg mL^{-1}, Tetracycline hydrochloride 10 µg mL^{-1}, disodium Carbenicillin 100 µg mL^{-1}.

Minimal media for shaking flask experiments and the precultures for bioreactor experiments consisted of 4 gL^{-1} glucose, 3.2 gL^{-1} $NaH_2PO_4{\cdot}2H_2O$, 11.7 gL^{-1} K_2HPO_4, 8 gL^{-1} $(NH_4)_2SO_4$, 0.01 gL^{-1} thiamine hydrochloride and 0.2% (V/V) trace elements stock solution. Minimal media for batch cultivation in the bioreactor consisted of 13.4 gL^{-1} glucose, 1 gL^{-1} $NaH_2PO_4{\cdot}2H_2O$, 2.6 gL^{-1} K_2HPO_4, 9 g/l $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. In the experiments with strains carrying pJOE4056.2_tetA for GFP production 10 µgmL^{-1} Tetracycline hydrochloride and 1 gL^{-1} Rhamnose were supplemented. Towards the end of the batch phase about 100 µL of antifoaming agent Struktol J647 was added to prevent foaming upon glucose depletion. Minimal media for continuous chemostat cultivation in the bioreactor consisted of 13.14 gL^{-1} glucose, 1 gL^{-1} $NaH_2PO_4{\cdot}2H_2O$, 2.6 gL^{-1} K_2HPO_4, 9 gL^{-1} $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. In the experiments with strains carrying pJOE4056.2_tetA for GFP production 10 µgmL^{-1} Tetracycline hydrochloride and 1 gL^{-1} Rhamnose were supplemented. Throughout the chemostat phase 50 µLh^{-1} of antifoaming agent Struktol J647 were added continuously to the fermentation medium.

The composition of trace element stock solution was: 4.175 gL^{-1} $FeCl_3{\cdot}6H_2O$, 0.045 gL^{-1} $ZnSO_4{\cdot}7H_2O$, 0.025 gL^{-1} $MnSO_4{\cdot}H_2O$, 0.4 gL^{-1} $CuSO_4{\cdot}5H_2O$, 0.045 gL^{-1} $CoCl_2{\cdot}$ $6H_2O$, 2.2 gL^{-1} $CaCl_2{\cdot}2H_2O$, 50 gL^{-1} $MgSO_4{\cdot}7H_2O$ and 55 gL^{-1} sodium citrate. Stock solutions of salts, trace elements and sugars were autoclaved separately, and stock solutions of thiamine hydrochloride and the antibiotics were filter sterilized and stored at 4°C. All compounds were combined just before the experiments to prevent potential aging of media.

PBS-MgCa for the measurement of eGFP fluorescence and flow cytometry analysis contained 8 gL^{-1} NaCl, 0.2 gL^{-1} KCl, 1.44 gL^{-1} Na_2HPO_4, 0.24 gL^{-1} KH_2PO4, 1 mM $MgSO_4$ and 0.1 mM $CaCl_2$. Prior to use PBS-MgCa was filtered with a sterile filter (pore size < 0.2 µm) to reduce particle load (Tomasek et al., 2018).

TABLE 4.1: Bacterial Strains used in this study

Strain	Genotype/Strain Information	Reference/Source
E. coli K-12 MG1655	F^-, λ^-, *ilv*G^-, *rfb*-50, *rph*-1 (‘wild type’ strain, abbrev. WT)	(Michalowski et al., 2017)
E. coli DH5α λ*pir*	*supE*44, Δ*lac*U169 (Φ80lacZΔM15), *recA*1, *endA*1, *hsdR*17, *thi*-1, *gyrA*96, *relA*1, λpir phage lysogen	(Michalowski et al., 2017)
E. coli DH10B pSIM5	F^- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ^- *rps*L(Str^R) *nup*G	(Datta et al., 2006)
T-SACK	W3110 *araD*<>*tetA-sacB-amp fliC*<>*cat argG*::*Tn*5	(Li, Thomason, et al., 2013)
E. coli CD101	MG1655 Δ*flk*	This study
E. coli CD201	MG1655 Δ*flk* Δ*fliA*	This study
E. coli CD202	MG1655 Δ*flk* Δ*fliA* Δ*fliC*	This study
E. coli CD203	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL*	This study
E. coli CD204	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR*	This study
E. coli CD205	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA*	This study
E. coli RM206	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD*	This study

E. coli RM207	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA*	This study
E. coli RM208	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR*	This study
E. coli RM209	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpTCBA*	This study
E. coli RM210	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpTCBA* Δ*yeeL*	This study
E. coli RM214	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpTCBA* Δ*yeeL* Δ*flxA*	This study
E.coli BW3110 pJOE4056.2	W3110 *rhaB*$^{-}$	(Wegerer et al., 2008)
E. coli DH5α λpir pJOE4056.2_tetA	*supE*44, Δ*lac*U169 (Φ80lacZΔM15), *recA*1, *endA*1, *hsdR*17, *thi*-1, *gyrA*96, *relA*1, λpir phage lysogen	This study

E. coli K-12 MG1655 rhaB$^-$	F$^-$, λ^-, *ilvG*$^-$, *rfb*-50, *rph*-1, *rhaB*$^-$	This study
E. coli RM214 rhaB$^-$	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpT* Δ*yeeL* Δ*flxA rhaB*$^-$	This study
E. coli K-12 MG1655 rhaB$^-$ pJOE4056.2_tetA	F$^-$, λ^-, *ilvG*$^-$, *rfb*-50, *rph*-1, *rhaB*$^-$	This study
E. coli RM214 rhaB$^-$ pJOE4056.2_tetA	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpT* Δ*yeeL* Δ*flxA rhaB*$^-$	This study

4.3.2 Construction of Deletion Strains

Chromosomal modifications were conducted using recombineering methods that have been comprehensively described and reviewed previously (Murphy, 2016). The *tetA-sacB* cassette and lambda recombineering functions provided by pSIM5 were used to perform chromosomal modifications with base-pair precision (Datta et al., 2006; Li, Thomason, et al., 2013). Deletions of single genes were designed to span the coding sequence only and deletions of operons or larger genomic regions were designed to begin with the coding sequence of the first gene and end with the coding sequence of the final gene. All deletions were verified by sequencing. Table B.1 contains an annotated list of primers used in this study and Supplementary information S2 in appendix B a more detailed description of the recombineering method used.

4.3.3 Construction of GFP production strains

The protein expression system used for the bioreactor fermentations closely resembles previously described systems based on pJOE4056.2 (Wegerer et al., 2008; Wilms et al.,

2001). For additional stability, the *bla* resistance cassette from plasmid pJOE4056.2 was exchanged for a *tetA* resistance cassette yielding pJOE4056.2_tetA to enable continuous selective pressure under the conditions of a chemostat. Use of pJOE4056.2_tetA requires induction with the rare sugar rhamnose at low glucose concentrations. Prior to plasmid transformation, we thus inactivated the chromosomal copy of *rhaB* encoding rhamnulokinase in *E. coli* MG1655 and *E. coli* RM214 to yield $rhaB^{-}$ strains incapable of utilizing the rare sugar rhamnose. Supplementary information S2 in appendix B contains a more detailed description of the procedure.

4.3.4 Shaking Flask Cultivations

For growth experiments glycerol stock cultures strains were streaked on 2xTY agar plates and incubated overnight at 37 °C. For precultures, a single colony was picked to inoculate 15 ml minimal medium in a 50 ml baffled shaking flask and incubated at 37 °C on an orbital shaker set to 130 rpm overnight. On the following morning, an inoculum of the preculture was transferred into 50 ml minimal medium in a 500 ml baffled shaking flask to reach a starting OD of 0.2 and the culture incubated at 37 °C on an orbital shaker set to 130 rpm. Samples were drawn hourly using a fixed needle reaching through the attached cotton plug and a syringe. In all shaking flask experiments the wild type strain *E. coli* MG1655 was cultivated in parallel as a reference and data collected from other strains was normalized to this reference data.

4.3.5 Bioreactor Setup

Bioreactor fermentations were carried out in a two-compartment scale-down reactor. The primary reactor was a stirred tank reactor, and a plug flow reactor was used as the secondary compartment mimicking a starvation zone. The plug flow reactor was connected to the stirred tank reactor only after establishment and sampling of a steady state in the chemostat phase. The technical setup has been characterized previously and includes the modifications described by Ankenbauer et al. (2020). A schematic overview of the two-compartment reactor is shown in figure 4.1 and Supplementary Information S2 in appendix B contains a comprehensive description of the setup.

4.3.6 Preculture, Batch Cultivation and Continuous Cultivation

100 µL of glycerol stock seed culture were directly used to inoculate 300 mL of preculture minimal media in a 3 L baffled shaking flasks and incubated at 37 °C on an orbital

shaker set to 130 rpm overnight. In the next morning 160 mL of preculture were used to inoculate the bioreactor. The total volume in the bioreactor was 1.6 L after inoculation. Batch fermentation in the bioreactor ensued at 37 °C. Upon depletion of glucose, indicated by a sharp increase in dissolved oxygen tension, feed and harvest lines were connected, the reactor refilled with feed medium to 1.6 L broth and a constant feed/harvest rate established. For the GFP production experiments with strains carrying pJOE4056.2_tetA the feed rate was set to 5.33 mL min^{-1} corresponding to a dilution rate of 0.2 h^{-1}. For the bioreactor cultivations aimed at investigating genomic stability and determining the maintenance coefficient of *E. coli* MG1655 und *E. coli* RM214 the batch phase was shortened and feed rates were set to 8.00 mL min^{-1}, 5.33 mL min^{-1}, 2.67 mL min^{-1} or 1.33 mL min^{-1} corresponding to dilution rates of 0.3 h^{-1}, 0.2 h^{-1}, 0.1 h^{-1} or 0.05 h^{-1}. After cultivation for at least five volumetric residence times a reference sample was taken. The plug-flow reactor was then connected to the primary reactor via a diaphragm metering pump effectively circulating about one-quarter (380 mL) of the total fermentation broth from the primary reactor through the plug-flow reactor and back into the stirred tank reactor. In the following five to six volumetric residence times samples were taken at predefined time points from the STR and the five PFR ports. Afterwards the fermentation was aborted, and the actual final broth volume measured. This value was used for all volumetric calculations during data analysis.

4.3.7 Determination of Optical Density and Biomass dry weight

Optical density of fermentation broth appropriately diluted with 0.9 NaCl from the primary reactor was measured in triplicates at 600 nm on a spectrophotometer (Amersham Biosciences/GE Healthcare, Amersham, United Kingdom). For the measurement of biomass dry weight quadruplicates of 5 mL of broth were centrifuged in weighted glass tubes at 2500 g and 4 °C for 7.5 min. Supernatant was immediately decanted and the pellet washed by resuspending in 5 ml of freshly prepared 150 mM NH_4HCO_3 held at 4°C. The suspension was centrifuged again, and the washing repeated once. After a final centrifugation, remaining liquid was decanted carefully, the pellet dried at 105 °C and glass tubes containing dried pellets weighted again.

4.3.8 Determination of Acetic acid, Ammonium and Glucose concentrations in fermentation supernatant and feed

5 mL of biosuspension was directly sampled into a syringe connected to a single-use 0.45 µm sterile filter and immediately filtered. The clear supernatant was flash frozen in liquid nitrogen and stored at -70 °C until analysis. Glucose concentration was determined by D-Glucose UV-Test Kit (R-Biopharm, Darmstadt, Germany) and acetic acid concentration by Acetic acid UV-Test Kit (R-Biopharm, Darmstadt, Germany). Ammonium concentration was determined by Ammonium cuvette test LCK 303 or LCK 304 (Hach Lange, Düsseldorf, Germany). At the end of the cultivation feed samples were taken directly from the feed line, flash frozen in liquid nitrogen and processed as described.

4.3.9 Analysis of Total Carbon, Inorganic Carbon and Biomass Composition

For total carbon and inorganic carbon analysis 0.5 mL biosuspension sample were mixed with 50 µL of 5 M KOH to prevent loss of dissolved carbonate. The suspension was then diluted 1:20 with demineralized water, flash frozen in liquid nitrogen and stored at -70 °C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) to yield the total concentration of carbon and inorganic carbon in the fermenter effluent stream.

To determine biomass composition 1.0 ml of biosuspension was centrifuged at 4 °C and 14000 rpm (20817 g) for 3 min. The supernatant was discarded, the pellet resuspended in 1.0 mL of 0.9% NaCl solution and centrifuged again. The pellet was resuspendend in 5 ml 0.9% NaCl, flash frozen in liquid nitrogen and stored at -70 °C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) and the carbon and nitrogen content of the biomass calculated from these values.

4.3.10 Measurement of Nucleotides

2 mL of biosuspension was sampled directly into 0.5 mL of precooled (< -20°C) quenching solution and incubated at 6 °C on a shaker for 15 min. Quenching solution consisted of 80 µM EDTA dissolved in 35% (V/V) perchloric acid. 500 µL 1 M K_2HPO_4 was added and the sample briefly vortexed. 550 µl 5 M KOH was added and the sample vortexed again. To remove precipitating potassium perchlorate samples were then centrifuged at 4 °C and 7830 rpm (7197 g) for 5 min. 1.5 mL of supernatant was carefully transferred to new tubes, flash frozen in liquid nitrogen and stored at -70 °C. Prior to analysis samples

were thawed, centrifuged for 10 min at 4°C and 7197 g, 1 mL of supernatant transferred to new tubes and their pH adjusted to 6.95 - 7.05 with 5 M KOH or 35% (V/V) perchloric acid. Samples were centrifuged again for 30 min at 4°C at 18000 g to remove potassium perchlorate precipitate from neutralization. 500 µL of supernatant were then transferred into RotiSpin Mini 3 kDa MWCO tubes and centrifuged again for 30 min at 4°C at 18000 g. HPLC analysis was carried out as described previously (Löffler, Simen, Jäger, et al., 2016b).

4.3.11 Measurement of eGFP Fluorescence

Freshly sampled biosuspension was flash-frozen in liquid nitrogen and stored at -70°C until analysis. On the day of analysis all samples were thawed and diluted 1:100 with ice-cold PBS-MgCa. 200 µL of diluted sample were transferred into a black 96 well-plate with transparent bottom and lid and the fluorescence (excitation 485 nm, emission 535 nm) quantified in a SLT SpectraFluor plate-reader (Tecan, Switzerland). The measured fluorescence values were then converted into absolute eGFP concentrations using a calibration curve recorded with purified protein (Supplementary Information S2 in appendix B).

4.3.12 Flow Cytometry Analysis

Freshly sampled biosuspension was diluted with PBS-MgCa to yield an OD of approximately 0.04. Diluted biosuspension was passed through a 30 µM CellTrics® filter to reduce particle content and analyzed in a BD Accuri^{TM} C6 Plus Flow Cytometer. The excitation laser had a wavelength of 488 nm and a 533/30 nm emission filter was used to capture GFP fluorescence. Particle signals with a forward scatter height (FSC-H) signal less than 2500 were ignored and 250000 events collected. Events with an eGFP area signal less than 10 were excluded from the analysis to remove dust and cell debris, usually resulting in 235000 - 249000 remaining events. Cells from events with an eGFP area less than 2000 were defined to form the non-producing population, while cells from events with an eGFP area equal or greater than 2000 were defined to form the producing population. Histograms of all samples can be found in Supplementary Figure B.3 and B.4 in appendix B.

4.3.13 Genomic DNA Sequencing

1 mL of biosuspension was sampled, flash-frozen in liquid nitrogen and stored at -70 °C. On the day of extraction samples were thawn and total DNA extracted with DNeasy Blood and Tissue Kit (Qiagen). Isolated DNA was shipped to and sequenced by the commercial sequencing partner Eurofins Genomics resulting in approximately 5 to 6 million paired end 150 bp reads per sample. Data was delivered as fastqc files and assembly of the reads conducted with Unicycler 0.4.8 with the following settings: min contig length 300 bp, min contig coverage 5 (Wick et al., 2017). The obtained contigs were processed with Mauve version 20150226 build 10 using the reference sequence NC_0000913.3 from the NCBI database (Darling et al., 2004). Finally, small nucleotide polymorphisms were detected using snippy (https://github.com/tseemann/snippy) and the output manually examined using Geneious Prime 2020.2.3 (https://www.geneious.com). Supplementary data S8 in appendix B contains lists of all SNPs found.

4.3.14 RT-qPCR

1.5 mL of freshly drawn biosuspension were immediately flash frozen in liquid nitrogen and stored at – 70 °C. Frozen liquid cell suspensions were thawn on ice and 200 µL each were transferred into bead bashing tubes prefilled with 700 µL Lysis buffer. Cells were disrupted with a Precellys® homogenisator for 2 x 20 s.RNA was extracted using the Quick-RNA Fungal/Bacterial Kit (Zymo Research) following the manufacturer´s instructions. The RNA concentrations were measured by Nanodrop. 10 µg RNA each was treated with 2 units TURBO DNase (Thermo Fisher Scientific) in 50 µL reactions for 60 min, with additional 2 units enzyme after 20 and 40 min, respectively. RNA from the DNase reactions were purified with Zymo Clean & Concentrator™-5 (Zymo Research) according to the manufacturer´s protocol and were then measured by Nanodrop. cDNA synthesis with SuperScript® IV reverse transcriptase (Invitrogen) was carried out according to the protocol for random hexamers as primers. 1 µg RNA was used as starting input for 20 µL reactions, but no RNase inhibitor was added. A no reverse transcriptase control was included. For the qPCR reactions, the cDNA reaction mixes were diluted with 100 µl nuclease free water. 2 µL from all cDNA reactions were pooled together and a dilution series was prepared (1, 1:10, 1:100, 1:1000) for determination of PCR efficiency for each primer pair during each PCR run. For 15 µL reactions 7.5 µL ORA™ qPCR Green ROX L Mix (highQu), 0.4 µL forward primer, 0.4 µl reverse primer (f.c. 266 nM, each), 4.7 µL H2O and 2 µL of diluted cDNA reactions were mixed. eGFP was amplified using primers eGFP2-forward

and eGFP2-reverse (amplicon length: 248 bp), for cysG primers cysG_housekeeping_fwd and cysG_housekeeping_reverse (amplicon length: 197 bp) were used. All reactions were performed as triplicates. Reactions were carried out on a Biorad CFX96 in 96 well plates. Program parameters were 95 °C, 3 min; 39x (95 °C, 5 sec; 59 °C, 15 sec; 72 °C, 15 sec); 65 °C to 95 °C (0,5 °C increment). Data was analyzed with Biorad CFX Manager 3.1. Relative expression of eGFP to cysG was calculated from the cq numbers measured by the instrument adjusted for amplification efficiency. Relative expressions from time points STR PFR 25 h and STR PFR 28 h were normalized to the corresponding STR sample.

4.4 Results

4.4.1 Engineering of *E. coli* deletion strains

Our primary goal was to engineer a series of deletion strains based on *E. coli* MG1655 with physiological advantages under heterogeneous conditions with nutrient depleted zones. Strains would ultimately be assayed in a scale-down reactor consisting of a primary stirred tank reactor (STR) and a secondary plug-flow reactor (PFR) mimicking a starvation zone (figure 4.1).

We began with defining criteria for the choice of handpicked deletion targets: First, only genes that cause relevant metabolic burden in the context of a large-scale bioprocess should be chosen. We thus based our choice of targets primarily on the list of genes with high add-on maintenance under repeated transient starvation published by Löffler, Simen, Müller, et al. (2017) and selected genes with an estimated add-on maintenance > 0.05 %. Except for *fliC* none of the chosen genes had an estimated maintenance add on > 1 %, so we expected very little contribution of most single deletions. It was thus clear that multiple deletions would be necessary to achieve reasonably measurable effects. To maximize potential savings, we removed the entire operon if a candidate gene was part of a functionally connected operon. Second, any deletion must not be detrimental to basic growth parameters in glucose minimal medium. In the past, *E. coli* deletion strain series such as the MDS or the MGF series, had suffered from biological fitness losses (Karcagi et al., 2016; Kurokawa et al., 2016). Learning from these studies, any genes involved in primary carbon metabolism or basic cellular functions were outright excluded and we aimed for a highly selective approach with a strictly limited scope. Third, global regulatory programs must be left intact to avoid potential side effects. This included the general stress response, SOS responses and the stringent response. The stringent response had previously been identified as the major repeatedly induced regulatory program but

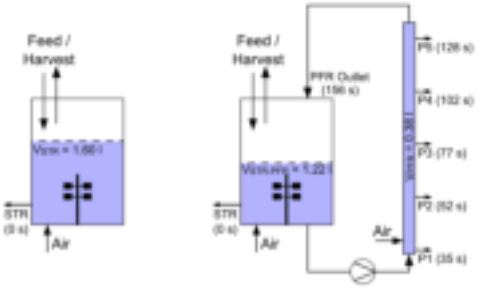


FIGURE 4.1: Reactor setup. The primary reactor was a standard laboratory reactor operated as a fully aerobic glucose limited chemostat at 37 °C (left scheme). For measurements of the well mixed STR reference state the entire biosuspension was in the primary reactor (V_{STR} = 1.60 L). Scale down conditions were installed by connecting a secondary plug flow reactor (PFR). An active pump then constantly circulated V_{PFR} = 0.38 L fermentation broth between STR and PFR reducing the volume fraction in the STR to $V_{STR-PFR}$ = 1.22 L (right scheme). Labels STR and P1 to P5 designate sampling ports with the respective average residence time of biosuspension after leaving the STR. Fermentations were carried out in two phases each lasting for at least five volumetric residence times: First, a homogeneous STR reference state was established, followed by a subsequent heterogeneous STR-PFR phase.

strains with modulated ppGpp availability already exist and have dampened regulatory patterns in nutrient-limited conditions (Michalowski et al., 2017; Ziegler et al., 2020). In this study, our goal was to work orthogonally to cellular regulation.

With these criteria in mind we developed a set of planned deletions containing most parts of the flagellar apparatus, the chemotaxis systems, and multiple other handpicked genes(with addon to maintenance > 0.05 %): *cspD*, *aldA*, *flxA*. CspD is a toxin of dispensable function, AldA is irrelevant in glucose-limited medium as its essential function is complemented by PrpC and FlxA a protein from the Qin prophage. All of these genes are non-essential (Baba et al., 2006). Using lambda recombineering with the *tetA-sacB* cassette we sequentially engineered the strains starting from *E. coli* MG1655 until completion of the final strain *E. coli* RM214 (table 4.1). We assayed any new deletion strain from the series for its basic growth parameters in shaking flask fermentations cultivating *E. coli* MG1655 as a benchmark in parallel. None of the deletion strains had major advantages or deficits in maximum specific growth rate or biomass yield in glucose minimal medium affirming our choice of deletion targets (figure 4.2).

Conducting the genomic deletions required a high number of total passages until *E. coli* RM214 was completed. We sequenced both the genome of *E. coli* MG1655 and *E.*

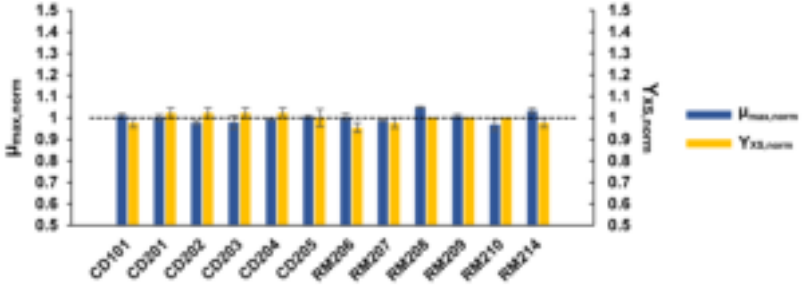


FIGURE 4.2: Basic growth parameters of deletion strains. Deletion strains CD101 to RM214 were cultivated in minimal glucose medium in shaking flask fermentations. The maximum specific growth rate (blue) and biomass yield (yellow) were determined. The parent strain E. coli MG1655 was cultivated in parallel, and all data collected normalized to its growth parameters. Error bars indicate SEM (n = 3). The dashed line is a visual aid indicating reference values of 1.

coli RM214 and identified no problematic mutations (Supplementary information S2 in appendix B, Supplementary Data S9). As we expected little impact of single deletions, we decided to focus our characterization only on the final strain of the series, *E. coli* RM214, and compared it to its parent wild-type strain *E. coli* MG1655.

4.4.2 Maintenance coefficient and genomic stability in scale-down fermentations

To test the initial hypothesis of a reduced maintenance coefficient in heterogeneous conditions and unravel potential benefits of *E. coli* RM214, we cultivated *E. coli* MG1655 and *E. coli* RM214 in two-compartment scale-down fermentations. Continuous chemostat cultivations with two phases were used to enable accurate assessment of fermentation parameters. In the first phase, strains were cultivated in standard well-mixed conditions employing only a STR (figure 4.1, left scheme). After five volumetric residence times this reference state was sampled and the secondary PFR compartment connected to the STR. A diaphragm metering pump then continuously circulated about one-fourth of the fermentation broth from the STR through the PFR and back into the STR. As feeding occurred only in the STR, the PFR simulated repeated passages of fractions of the population through a starvation zone (figure 4.1, right scheme). After continued cultivation for another five volumetric residence times the new STR-PFR steady state was sampled. Therefore, the total process time always exceeded ten volumetric residence times.

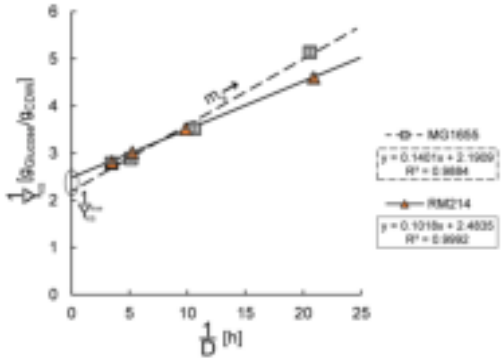


FIGURE 4.3: Determination of maintenance coefficients under heterogeneous STR-PFR conditions. *E. coli* MG1655 (grey squares, dashed line) and *E. coli* RM214 (orange triangles, solid line) were cultivated in the STR-PFR system (glucose limited chemostat, D = 0.05 h^{-1}, 0.1 h^{-1}, 0.2 h^{-1}, 0.3 h^{-1}). Maintenance coefficients m_s (slope) and true biomass yields Y_{XS}^{true} (intersection) were determined from the linear regression of data points. The difference in maintenance coefficients is statistically significant (Δm_s = - 0.038 $g_{Glucose}$ g_{CDW} $^{-1}$ h^{-1}, $p < 0.05$). Error bars indicate technical standard deviation.

We cultivated *E. coli* MG1655 and *E. coli* RM214 at four different dilution rates (0.05 h^{-1}, 0.1 h^{-1}, 0.2 h^{-1}, 0.3 h^{-1}) each. We measured biomass concentrations in the well mixed STR reference state and during the heterogeneous STR-PFR phase. Biomass yield on substrate was in general similar for both strains but *E. coli* RM214 had a slightly increased biomass yield on substrate, especially under STR-PFR conditions and at D = 0.05 h^{-1}. We estimated Pirt's maintenance coefficient m_s of both strains by linear regression of Y_{XS}^{-1} vs D^{-1} (figure 4.3). We found no statistically significant differences under well mixed STR conditions but the maintenance coefficient of *E. coli* RM214 was significantly lower than that of *E. coli* MG1655 under STR-PFR conditions (Δm_s = - 0.038 $g_{Glucose}$ g_{CDW} $^{-1}$ h^{-1}, $p < 0.05$). Differences in the true biomass yield $Y_{XS}{}^{true}$ were not significant under any conditions ($p > 0.05$). The results confirm the effectiveness of our deletion strategy specifically for the targeted environment.

We sequenced the strains' genomes from the STR-PFR samples from all fermentations to investigate potential genomic instability due to the long fermentation time (> 200 h at D = 0.05 h^{-1}). In *E. coli* MG1655, we found SNPs in *insH5* in samples from all dilution rates but no other mutations. We also found SNPs in *insH5* in all samples from *E. coli* RM214 and additional mutations in *ycfk* and *stfE* of the inactive e14 prophage (Supplementary

Data S9 in appendix B). Apart from these minor alterations, the strains were remarkably stable. They showed no accumulation of mutations in any regulatory genes or genes involved in central metabolism confirming that the engineered deletions bestowed the reduced maintenance coefficient to *E. coli* RM214.

4.4.3 Construction of eGFP production strains

Based on these encouraging findings we hypothesized that *E. coli* RM214 should better withstand the stressful conditions of an exemplary heterogeneous production scenario than its ancestor strain *E. coli* MG1655. We chose to produce eGFP as an easily measurable proxy for industrially relevant intracellularly accumulated proteins such as insulin varieties or other biopharmaceuticals commonly produced in *E. coli* (Baeshen, Al-Hejin, et al., 2015; Baeshen, Baeshen, et al., 2014).

A suitable expression system to produce proteins in glucose-limited fermentations is the rhamnose-inducible expression system from pJOE4056.2 (Wegerer et al., 2008). Expression from the rhamnose promoter occurs in the presence of non-toxic rhamnose and is enhanced by low levels of glucose sensed by cAMP-CRP signaling. However, the use of rhamnose as a stable inducer requires the absence of rhamnose catabolism (Wilms et al., 2001). We therefore inactivated the chromosomal copy of *rhaB* by replacing the original gene in *E. coli* MG1655 and *E. coli* RM214 with an inactive frameshift copy from *E. coli* BW3110 by recombineering with the *tetA-sacB* cassette. The resulting strains were termed *E. coli* MG1655 rhaB$^-$ and *E. coli* RM214 rhaB$^-$. The absence of rhamnose catabolism was additionally confirmed by streaking the strains on 2xTY ph indicator agar plates containing rhamnose. *E. coli* MG1655 rhaB$^-$ and *E. coli* RM214 rhaB$^-$ formed white colonies meaning that no acidification of the medium caused by rhamnose degradation occurred.

We then exchanged the *bla* resistance gene from pJOE4056.2 for the *tetA* resistance gene from *E. coli* T-SACK generating pJOE4056.2_tetA (Supplementary figure B.1). TetA is a tetracycline exporter and thus enables continuous selective pressure in the presence of tetracycline during prolonged cultivations. Transformation of the *rhaB*$^-$ strains with pJOE4056.2_tetA yielded *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA (table 4.1).

4.4.4 Scale-down fermentations with eGFP production

E. coli MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA were then fermented in quadruplicates each in the STR-PFR scale-down reactor in continuous

chemostat cultivations at a dilution rate of D = 0.2 h^{-1}. Heterogeneities were introduced by using the two-compartment STR-PFR reactor in the same setting as described above (figure 4.1). Again, this included a well-mixed STR only chemostat phase, and a subsequent STR-PFR chemostat phase to enable direct observation of the influence of the nutrient-limited zone.

Under well-mixed STR conditions, we observed no substantial differences between the fermentations of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. They reached comparable cell dry weight and eGFP yield on glucose (figure 4.4). In fact, the strains had virtually identical fermentation and production parameters in any parameter measured (Table 4.2). The primary product eGFP formed a considerable fraction of the total biomass and we detected only trace amounts of acetate byproduct as expected for glucose-limited fermentations. We also determined the proportion of cells with high eGFP content by flow cytometry and found these to be practically identical for both strains in the STR reference steady-state (figure 4.4). As *E. coli* RM214 was specifically engineered to have advantageous traits in heterogenous fermentations including starvation zones these findings were not surprising and instead proved that our genomic deletions do not interfere with the basic fermentation traits of *E. coli* K-12 strains.

Upon connecting the PFR the process performance of both strains started to decline, but this phenomenon occurred remarkably slower and much less pronounced in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA than in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. Five hours after connection of the PFR both strains still had similar fractions of producing cells and reached comparable biomass concentration. However, first differences in cellular eGFP content and product yield already became apparent. Over the remaining process time production parameters increasingly diverged. After 28 h of STR-PFR continuous cultivation we observed a 43% higher product yield in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA than in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (ΔY_{PS} = 13 $mg_{eGFP} {g_{Glucose}}^{-1}$, two-tailed t-test, $p < 0.05$). Concomitantly, the proportion of actively producing cells shrank rapidly in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and more slowly in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. Instead, biomass concentration increased in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA indicating a shift from production to biomass formation (Supplementary Figure B.4 A). Noteworthy, we found a linear correlation describing the tradeoff between eGFP production and biomass formation using data from both *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA (Supplementary Figure B.4 C and B.4 D). We suspected that the divergence may be caused by a reduced fraction of producing cells for *E.*

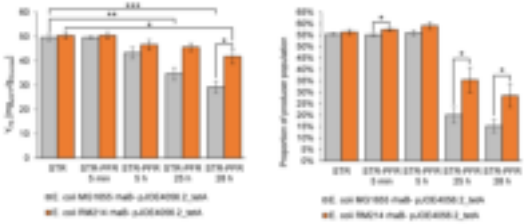


FIGURE 4.4: eGFP yield on substrate and proportion of cells with high eGFP content. *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (grey) and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA (orange) were cultivated in the STR-PFR system (glucose limited chemostat, D = 0.2 h^{-1}). Samples were collected from the primary vessel. Error bars indicate SEM (n = 4), statistical indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. **Left:** eGFP yield on substrate declines for both strains after PFR connection. Simultaneously, the difference between the strains gradually increases. Statistics: two tailed t tests comparing means of a single strain at later time points to the STR mean of the strain; and comparing the means of both strains at each time point to each other. **Right:** The proportion of cells with high eGFP content declines towards the end of the fermentation and is lower for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA than for *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. Statistics: one tailed t tests comparing the presumably lower mean of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA to that of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA at each time point.

coli MG1655 rhaB$^-$ pJOE4056.2_tetA compared to *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA and measured the fluorescence of individual cells by flow cytometry. Similar to the eGFP yield, the proportion of actively producing cells shrank rapidly in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and more slowly in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. At the final time point the fraction of producing cells was significantly higher for *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA than for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (one tailed t test, $p > 0.05$). To check whether differential expression of eGFP might be responsible for the reduction of eGFP yield in the heterogeneous conditions in general or for the differences between the two strains, we conducted RT qPCR using the housekeeping gene *cysG* as a reference. However, we found no clear indication for differential expression of eGFP towards the end of the fermentation or between the two strains (figure B.9 in appendix B).

After connection of the PFR we observed alterations in the respiratory parameters of both strains. Initially, cells reacted with a short spike of increased respiratory activity which then dropped rapidly in the following hour. The oxygen uptake rate Q_{O2} and the carbon dioxide formation rate Q_{CO2} recovered over the next two volumetric residence times and then slowly drifted towards new steady states but never reached the initial STR

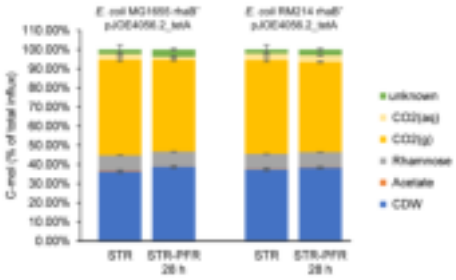


FIGURE 4.5: Carbon Balance. *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (grey) and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA (orange) were cultivated in the STR-PFR system (glucose limited chemostat, D = 0.2 h^{-1}). Columns show efflux fractions of individual substances. Error bars indicate SEM (n = 4). For raw data see Supplementary Tables B.3 and B.4.

only values (Supplementary Figure B.4 and B.4). We calculated total carbon balances but the deviations in the respiratory rates caused only minor redistributions between the STR reference status and the STR-PFR 28 h sample (figure 4.5, Table B.3 and B.4). Apart from small gains in the biomass (CDW) fraction and small reductions in the carbon dioxide formation no major differences occurred. Declining productivity was hence accompanied by declining respiration and increased biomass formation. From all collected indications we conclude that the primary factor for loss of productivity was a restructuring of the biomass composition towards lower eGFP content (Figure B.4 B). This is supported by our observations using flow cytometry. The proportion of cells with high eGFP content dropped substantially in the late fermentation stages (figure 4.4). We presume that the reduced cellular eGFP content then led to lower metabolic burden and thus enabled slightly higher biomass yields. In all parameters measured, *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA proved to be more robust to the STR-PFR conditions and maintained productive for a longer period than *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. Since the only clearly different parameter between the two strains is the maintenance coefficient, we propose that *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA benefits from a small surplus of substrate that can be used to meet the high precursor and ATP demand of heterologous protein synthesis.

The energetic state of cells during cultivations, can be assessed by calculating the Adenylate Energy Charge (AEC) from measured nucleotide concentrations (Chapman, Fall, et al., 1971). Initially, in the well-mixed STR only phase, the concentration of all nucleotides and the AEC was comparable for both strains (Supplementary figure B.5). After connection of the PFR, we then simultaneously sampled cells from the STR and the five

ports along the primary axis of the PFR yielding a time-resolved profile of the AEC during PFR passage (figure 4.6). As expected during passage through a nutrient starvation zone, the AEC of cells dropped rapidly after leaving the STR and continued to decline towards a plateau. Shortly after PFR connection, pattern was highly similar for both strains (Figure 4.6, upper panel). After 25 h of cultivation under scale down conditions, the AEC of both strains in the STR and at all sampling ports of the PFR was higher than before (figure 4.6, lower panel). Here, differences between the strains also became apparent as the AEC of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA was higher than that of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA at all sampling points. We then compared the AEC of samples drawn from the primary vessel at different time points to unravel long term effects of the heterogeneous conditions. Both strains individually showed statistically significant increases in the AEC between time points STR and STR PFR 25 h (two tailed t tests, $p < 0.05$; see Supplementary Table B.2). In fact, the AEC of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA sampled from the primary fermentation vessel (figure 4.6, 0 s) at time point STR-PFR 25 h was the highest recorded value from all samples indicating that the strain was possibly trying to adapt to the unfavorable conditions. The coincidence with its reduced productivity and slightly increased biomass yield at the late fermentation stages points towards the preservation of cellular energy at the expense of heterologous protein productivity. The data from *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA indicates a similar but less pronounced trend. Comparing the two strains to each other reveals a marginally significant difference ($p = 0.077$; see Supplementary Table B.2) of the AEC values measured in samples from the STR at STR PFR 25 h which is reflected by the generally slightly lower AEC values of the deletion strain at this time point (figure 4.6, lower panel). It is noteworthy that the total AxP levels of both strains were comparable for all samples and only the distribution among ATP, ADP and AMP varied (Supplementary Figure B.5).

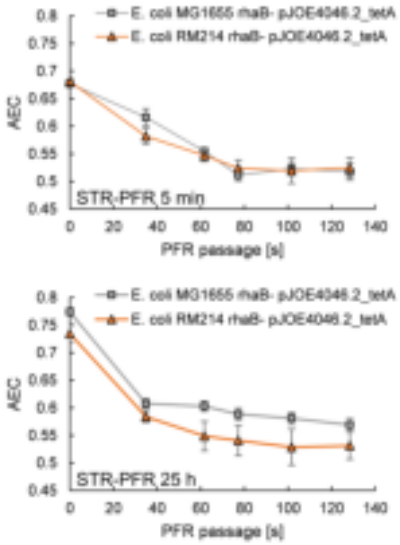


FIGURE 4.6: Adenylate Energy Charge in the STR and during PFR passage. *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA were cultivated in the STR-PFR system (glucose limited chemostat, D = 0.2 h^{-1}). The adenylate energy charge of cultures was determined shortly after PFR connection (STR-PFR 5 min, upper panel) and after five volumetric residence times (STR-PFR 25 h, lower panel). Samples were drawn from the primary reactor (0 s) and the five sampling ports along the axis of the PFR (35 s, 52 s, 77 s, 102 s, 128 s). Error bars indicate SEM (n = 4).

TABLE 4.2: Fermentation parameters of the eGFP production chemostat processes

		***E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA**		***E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA**	
	Unit	**STR**	**STR-PFR 28 h**	**STR**	**STR-PFR 28 h**
c_x	$\left[\frac{gCDW}{l}\right]$	4.02 ± 0.066[a]	4.28 ± 0.024	4.02 ± 0.053	4.13 ± 0.024
c_P	$\left[\frac{mg_{eGFP}}{l}\right]$	630 ± 20	370 ± 32	650 ± 11	540 ± 41
Y_{XS}	$\left[\frac{gCDW}{g_{Glc}}\right]$	0.315 ± 0.0048	0.3360 ± 0.00075	0.314 ± 0.0032	0.322 ± 0.0034
Y_{PS}	$\left[\frac{mg_{eGFP}}{g_{Glc}}\right]$	50 ± 1.3	29 ± 2.4	50 ± 1.2	42 ± 3.0
$c_{Ac,STR}$	$\left[\frac{g_{Ac}}{l}\right]$	0.14 ± 0.040	0.017 ± 0.0072	0.010 ± 0.0061	0.07 ± 0.030
$c_{NH_4^+}$	$\left[\frac{g_{NH_4^+}}{l}\right]$	1.64 ± 0.025	1.6 ± 0.12	1.57 ± 0.058	1.52 ± 0.033
q_S	$\left[\frac{g_{Glc}}{g_{CDW}h}\right]$	0.64 ± 0.013	0.60 ± 0.012	0.62 ± 0.020	0.60 ± 0.018
q_P	$\left[\frac{mg_{eGFP}}{g_{CDW}h}\right]$	32 ± 1.4	18 ± 1.8	31 ± 1.2	25 ± 2.0
Q_{CO_2}	$\left[\frac{mmol_{CO_2}}{h}\right]$	73.5 ± 0.70	70.6 ± 0.54	72.6 ± 0.49	69.8 ± 0.64
Q_{O_2}	$\left[\frac{mmol_{O_2}}{h}\right]$	69.9 ± 0.90	67 ± 1.2	69.8 ± 0.90	65.6 ± 0.44
RQ	$\left[\frac{mol_{CO_2}}{mol_{O_2}}\right]$	1.05 ± 0.019	1.06 ± 0.013	1.04 ± 0.011	1.06 ± 0.010
eGFP content	$\left[\frac{w}{w}\%\right]$	15.8 ± 0.65	8.7 ± 0.73	16.1 ± 0.46	13 ± 1.0
Prod. population	[%]	55.4 ± 0.43	15 ± 2.8	56 ± 1.2	28 ± 5.1
D	$\left[h^{-1}\right]$	0.201 ± 0.0026		0.194 ± 0.0028	

[a]Errors indicate SEM (n = 4).

4.5 Discussion

In this study, we created a series of deletion strains lacking genes with high addon-to-maintenance under heterogeneous conditions with repeated starvation. The final strain

of the series, *E. coli* RM214 had a significantly lower maintenance coefficient than its parent *E. coli* MG1655 in an STR-PFR scale-down reactor. Moreover, *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA proved to be more robust to the influence of heterogeneities in an exemplary protein production scenario reaching a significantly higher product yield in the STR-PFR phase.

The core concept of our deletion approach was to remove genes that are wastefully expressed under transient starvation conditions. The expected individual contribution of each single gene was very low (Löffler, Simen, Jäger, et al., 2016b). The only remarkable exception was *fliC* whose expression alone was estimated to cause add on maintenance of 3.10%, by far exceeding the expected add on maintenance of 0.55% for the second in line *aldA* (Löffler, Simen, Jäger, et al., 2016b). Multiple other flagellar and chemotaxis genes were also candidates, so the removal of these systems formed a major fraction of the deletions conducted in the creation of *E. coli* RM214. As the goal of this study was to investigate the fundamental usefulness of the whole design approach and each individual deletion had likely little effect, we did not attempt to experimentally assess the individual contributions or potential interactions. We measured a relative difference of around 38% between the maintenance coefficients of *E. coli* MG1655 and *E. coli* RM214 under scale down conditions which exceeds the sum of all individual contributions from transcriptional and translational metabolic costs as estimated by Löffler, Simen, Jäger, et al. (2016b). The calculations by Löffler, Simen, Jäger, et al. (2016b) were thus either very conservative or the actual absence of the expressed proteins provided additional secondary benefits. This appears particularly likely regarding the absence of the motility system which could save proton motive force otherwise used for flagellar rotation. We can also confidently exclude the possibility that the slightly reduced genome of *E. coli* RM214 had a major impact due to reduced replication cost. The combined size of the deletions in *E. coli* RM214 was only about 50 kb, just slightly more than 1% of the *E. coli* K-12 genome, and the metabolic demand of DNA replication is per se very low (Stouthamer, 1973).

Key findings of our study are the reduced maintenance demand of *E. coli* RM214 compared to its parent strain and the slower product yield decline of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA under STR-PFR conditions. Both differences must be caused by the genotype of *E. coli* RM214 but do these findings relate to each other by a causative link or did we observe correlated phenomena? The overall eGFP transcript levels in the production scenario were fairly stable (Supplementary Information S10 in appendix B), so it appears unlikely that differential expression of eGFP from pJOE4056.2_tetA causes the different yields. The flow cytometry data indicate that microbial individuality may play

a role, but without a plausible mechanism we must assume that this is a correlated observation. Instead, we opine that a connecting mechanism can be drawn from the overall balances and the strains' cultivation parameters. Substrate consumed for maintenance demand is, by definition, not available for biomass formation. It is commonly assumed that it is fully converted into terminal products and the energy available from this conversion harnessed by the cells as ATP to meet their maintenance demand (Stouthamer and Bettenhaussen, 1973). Thus, a logical link between maintenance coefficient and protein yield exists: A reduced maintenance coefficient means that more substrate is available for biomass or product formation including potential secondary ATP costs that may arise from a high foreign protein content. A simple estimation allows us to test the quantitative feasibility of a causative relation by comparing the magnitude of the reduced maintenance coefficient to the difference in protein yield: Given that there were no significant differences in $Y_{XS,true}$ we can assume that Δm_s is entirely available for the additional production of eGFP in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. Using Y_{XS} from STR-PFR 28 h and an assumed protein content of roughly 65% (Taymaz-Nikerel, Borujeni, et al., 2010) the difference in maintenance demand could sustain an additional eGFP production rate of no more than $\Delta q_{p,ms}$ = 9.3 $mg_{eGFP}/g_{CDW}/h$. The experimental difference of $\Delta q_{p,ms}$ = 7 $mg_{eGFP}/g_{CDW}/h$ falls well within that range. A causative relation between the two observations is thus quantitatively feasible, and in our opinion likely. In this case up to 82% of the saved substrate due to lower ms could have been used for the formation of additional eGFP in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA.

The connection between maintenance demand, energy availability and eGFP production is also supported by the AEC data collected. We found a declining AEC during PFR passage for both *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA and *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA at all time points, which is similar to the pattern observed in a preceding study with non producing *E. coli* K 12 (Löffler, Simen, Jäger, et al., 2016b). However, it is important to note that we measured lower AEC values in the STR and a steeper decline in the PFR, putatively due to heterologous protein production. In the heterogeneous fermentation phase, when productivity declined, we measured increased AEC values, especially in samples of the less productive *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (Supplementary Information S4 in appendix B). The AEC is a measurement of the energetic state of cells and usually tightly balanced in a range between 0.7 and 0.9 (Chapman, Fall, et al., 1971). The activity of many cellular processes is connected to the AEC and a lower AEC is associated with the activation of catabolic enzymes to meet cellular energy demands (Atkinson,

1968). Substrate depletion generally causes a reduction of the AEC (Chapman and Atkinson, 1977). Conversely, reduced AEC values have been reported in conditions when cells experienced high anabolic demand or high secondary metabolic costs, for instance caused by cultivation at their maximum specific growth rate, or induction of motility (Lieder et al., 2015; Martinez-Garcia et al., 2014). Heterologous protein induction is known to cause increased ATP maintenance demand (Weber et al., 2002). We thus propose that the AEC values measured from samples drawn from the primary reactor at different time points of the fermentations can be explained by the ATP demands associated with eGFP productivity. It appears likely that the generally lower AEC measured in this study compared to data from non producing *E. coli* K 12 cultivated under similar conditions is caused by the production of eGFP. The significant increases of the AEC values of both strains towards the end of the fermentations are then a consequence of their diminishing eGFP productivity. This also explains the more pronounced AEC increase and concomitant eGFP yield decrease of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA compared to the deletion strain. The question then arises to what extent the lower maintenance coefficient of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA under scale down conditions influences the AEC values. From data collected in the maintenance study (figure 4.3) and the eGFP production fermentations we can roughly estimate the ATP demand for eGFP production of both strains at time point STR PFR 25 h (Supplementary Data S8 in appendix B). About 13% of the total ATP demand of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and 22% of the total ATP demand of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA can be attributed to eGFP production. Despite its lower maintenance coefficient the combined ATP demand for maintenance plus eGFP production of the highly productive *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA then still exceeds the respective values of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA which is reflected by its lower AEC at this time point.

A secondary observation made in this study was that loss of productivity in the STR-PFR condition was accompanied by a decline in the proportion of highly productive cells (figure 4.4). Microbial population heterogeneity is a subject of intense research (Binder et al., 2017) and our data provides no clear explanation why this shift occurs. Two things should be noted: First, the population heterogeneity for both *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA is of the bimodal kind (supplementary information S3 in appendix B) and the fractions of producing cells in the homogeneous STR cultivation phase are practically identical for both strains. Second, once the PFR is activated, we saw a decrease in the fraction of highly productive cells in all fermentations (figure 4.4, supplementary figure B.3 and B.4) but the decrease was faster

and more consistent for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. The overall level of population heterogeneity is generally high since only slightly more than half of all cells are strongly accumulating eGFP. We presume this is caused by the interplay of our expression system and the fermentation conditions. The regulation of rhamnose catabolism is autocatalytic and thus bimodality might be caused by a similar mechanism as in the case of expression from the arabinose promoter P_{BAD} (Khlebnikov et al., 2000). However, transcript measurements by RT qPCR did not lead to a concise pattern that would explain both the differences between the two strains and the declining eGFP yield of each individual strain over the course of the heterogeneous fermentation phase. Given the fairly stable expression of eGFP and the continuous selective advantage provided by *tetA*, it also appears unlikely that plasmid loss or mutations were the underlying cause. Moreover, the general stability of eGFP expression from pJOE4056.2 has been determined to be perfect for over 50 generations in earlier studies (Wegerer et al., 2008).

The deletion approach in this study differs from previous works because target selection was based on existing expression data and limited to candidates that imposed a high metabolic burden but were irrelevant under the specified conditions (Valgepea et al., 2015). Large scale genomic deletions, the contrary approach, have been conducted before in *E. coli* K-12, for instance as part of the construction of the MDS strains (Posfai et al., 2006). These strains had little benefits in standard protein production scenarios over their wild-type parent and were even inferior in basic process parameters, potentially caused by disrupted regulation (Karcagi et al., 2016; Sharma, Campbell, et al., 2007; Sharma, Blattner, et al., 2007). It needs to be emphasized that our deletion strategy only provided advantages in the specified conditions of a heterogeneous bioprocess with transient starvation as *E. coli* RM214 had no benefit compared to *E. coli* MG1655 under well-mixed conditions. In this regard, it also appears clear that the deletion targets chosen by us cannot be directly transferred to other hosts or conditions since the naturally evolved regulation might be divergent. This is exemplified by an interesting comparison of our results to existing data from *Pseudomonas putida*. The exposure of *P. putida* KT2440 to heterogeneous STR-PFR conditions led to an increased and potentially wasteful expression of *fliC* similarly as in the case of *E. coli* (Ankenbauer et al., 2020). However, the deletion of the flagellar apparatus in *P. putida* EM329 led to significant improvements of basic fermentation and protein production parameters already under well mixed conditions (Lieder et al., 2015; Martinez-Garcia et al., 2014). This demonstrates not only the diverging regulation of motility between different microbes, but also implies that a strain like *P. putida* EM329 might unintentionally display additional beneficial traits in heterogeneous fermentations

with starvation zones.

A premise of our study was to work orthogonally to global cellular regulation. The general stress response, stringent response and SOS responses were not modified as they provide important functions for cellular adaptation and some expression systems depend on intracellular signaling molecules of global regulatory circuits. Examples include not only the CRP cAMP dependent system used in this study but also novel adaptive expression systems that autoregulates cellular stress (Lo et al., 2016). Interestingly, in a former study, the rapid inactivation of *rpoS* in homogeneous chemostat cultivations was reported, which pointed to a large selective advantage of mutants (Notley-McRobb et al., 2002). In contrast, we did not find any mutations in genes involved in the stringent response or the general stress response for neither *E. coli* MG1655 nor *E. coli* RM214 at any growth rate. We conclude that under heterogeneous conditions the selective pressure on inactivating global regulatory programs is either very low or their activation may even be favorable for cellular viability which affirms our neutral approach to cellular regulation. However, this does not mean that modulating cellular regulation could not be beneficial for process or production parameters. Recently, several *E. coli* knock out strains lacking hand-picked genes that are connected to post-induction stress responses were presented (Sharma, Shukla, et al., 2020). These strains have advantageous traits for protein production which could be integrated in *E. coli* RM214.

Our study design focused on the influence of starvation zones on microbial culture performance. The carbon limited STR contained a substrate limited growth and production zone representing the bulk of large scale fermentations. The PFR served to introduce a transient starvation stimulus representative of repeated passages through a hunger zone as predicted to occur in large scale reactors (Haringa, Deshmukh, et al., 2017; Kuschel and Takors, 2020). Since our experimental setup was specifically chosen for the study of transient starvation, it does not capture the effects of other heterogeneities, in particular transient substrate excess. It is well known that close to the feeding point, substrate excess and concomitant oxygen limitation dominate the environment in large-scale fed-batch processes. *E. coli* typically reacts to such conditions with the production of solvents or small organic acids caused by overflow metabolism or anaerobic fermentation (Lara, Taymaz-Nikerel, et al., 2009). The formation of byproducts then results in process performance losses even if reuptake in zones with lower nutrient concentration is possible (Enfors et al., 2001; Neubauer, Åhman, et al., 1995). Since our deletion approach was only aimed at reducing the additional metabolic costs of transient starvation, *E. coli* RM214 probably

responds to glucose excess like other *E. coli* K 12 strains. In principle, copying the design approach to construct an *E. coli* strain with reduced additional maintenance in excess zones appears to be feasible as the transcriptional response of *E. coli* MG1655 to glucose excess is large and involves many potentially process irrelevant genes (Veit et al., 2007). However, the resulting genetic modifications would not reduce process performance loss from byproduct formation which is likely the dominating issue in substrate excess conditions. Various strategies to alleviate byproduct metabolism have been developed by other research groups, such as the use of alternative substrate transporters, knock-outs or the expression of recombinant *Vitreoscilla* hemoglobin (Eiteman et al., 2006; Pablos et al., 2014; De Anda et al., 2006). Given that our strain design approach avoids modifications to global regulation or central carbon metabolism, we are confident that it is compatible with these existing strategies and their combination could result in chassis strains for generally robust scale up. A limitation of our study originates from the focus on the model protein eGFP. However, recombinant protein production is frequently limited by the availability of cellular precursors and ATP, so it is not far fetched to expect similar effects with other protein products (Glick, 1995; Heyland et al., 2011). The reduced maintenance coefficient of E. coli RM214 should also be helpful to produce molecules formed in ATP intensive pathways such as terpenoids (Li and Wang, 2016; Ward et al., 2018). Potential advantages could also occur when the accumulation of toxic products causes increased ATP demand for product export, membrane maintenance or pH homeostasis (Sun, Zahir, et al., 2011; Tsukagoshi et al., 2000). On the other hand, it may be less helpful when the formation of a small molecule product is connected to net ATP synthesis. Glycolytic flux depends on the ATP requirements of cells and in such cases enforced ATP wasting can even increase the production rate (Koebmann et al., 2002; Boecker et al., 2019).

With the increasing knowledge about cellular metabolism and its interplay with the heterogeneous conditions of large-scale processes new possibilities to improve process performance arise. In a recent review of Wehrs et al. (2019) emphasized that strains should be engineered specifically for the demands of large-scale production (Wehrs et al., 2019). In this context, our series of deletion strains is the first step towards host strains robust against the repeated exposure to starvation zones.

4.6 Conclusion

Large-scale fermentations often suffer from process performance loss due to heterogeneous environments. *E. coli* RM214 was engineered to obtain a deletion strain with reduced maintenance and superior production properties in fermentations with starvation zones. Our study is the first that aimed to improve a microbe by repeated genomic deletions for enhanced robustness towards heterogeneous conditions. The exemplified application of *E. coli* RM214 for eGFP production demonstrates the cellular capaciy to exploit the mainenance advantage for preventing non-wanted performance loss in heterogeneously mixed indutrial production scenarios. Although only showcased for eGFP, the strain offers the capacity to serve as a platform for a variety of different products. Notably, this complements classical scale-up engineering and should be a highly valuable tool to prevent non-wanted performance of essential Titer-Rate-Yield values under industrial production conditions.

Chapter 5

Transcriptional Profiling of the Stringent Response Mutant Strain *E. coli* SR Reveals Enhanced Robustness to Large-Scale Conditions

Authors

Martin Ziegler (shared co-first authorship) – University of Stuttgart, Institute of Biochemical Engineering
Julia Zieringer (shared co-first authorship) – University of Stuttgart, Institute of Biochemical Engineering
Ralf Takors (corresponding author) – University of Stuttgart, Institute of Biochemical Engineering

5.1 Abstract

In large-scale fed-batch production processes microbes are exposed to heterogeneous substrate availability caused by long mixing times. *Escherichia coli*, the most common industrial host for recombinant protein production, reacts by recurring accumulation of the alarmone ppGpp and energetically wasteful transcriptional strategies. Here, we compare the regulatory responses of the stringent response mutant strain *E. coli* SR and its parent strain *E. coli* MG1655 to repeated nutrient starvation in a two-compartment scale-down reactor. Our data shows that *E. coli* SR can withstand these stress conditions without a ppGpp mediated stress response maintaining fully functional ammonium uptake and

biomass formation. Furthermore, *E. coli* SR exhibited a substantially reduced short-term transcriptional response compared to *E. coli* MG1655 (less than half as many differentially expressed genes). *E. coli* SR proceeded adaptation via more general SOS response pathways by initiating negative regulation of transcription, translation, and cell division. Our results show that locally induced stress responses propagating through the bioreactor do not result in cyclical induction and repression of genes in *E. coli* SR, but in a reduced and coordinated response, which makes it potentially suitable for large-scale production processes.

5.2 Introduction

Heterogeneities in large-scale fed-batch bioprocesses have long been recognized as a cause for process performance loss at industrial scale compared to homogeneous processes at lab scale (Bylund, Collet, et al., 1998). Due to physical, economical and engineering constraints the generation of gradients in large-scale reactors is inevitable. Hydrostatic pressure influences the solubility and transfer of gasses and the mixing time of large reactors can be orders of magnitude higher than that of laboratory reactors producing strong measurable chemical gradients (Delvigne, Destain, et al., 2006; Enfors et al., 2001; Larsson, Törnkvist, et al., 1996; Junker, 2004). Common consequences of spatial heterogeneities are loss of productivity, reduced biomass yield, increased byproduct formation and genetic or plasmid instability (Bylund, Castan, et al., 2000; Bylund, Collet, et al., 1998; George et al., 1993; Neubauer, Häggström, et al., 1995; Hopkins et al., 1987; Jonge et al., 2011). Reduced process performance is not limited to a single species but can be observed for many industrial workhorse organisms like *Escherichia coli*, *Saccharomyces cerevisiae*, *Penicillium chrysogenum* and *Bacillus subtilis* (George et al., 1993; Jonge et al., 2011; Junne et al., 2011; Larsson and Enfors, 1988).

Due to the enormous costs associated with using and maintaining large-scale equipment, few experiments in the context of academic research have been performed in industrial scale bioreactors (Bylund, Castan, et al., 2000; Bylund, Guillard, et al., 1999; Enfors et al., 2001). In consequence, researchers have relied on the use of computational fluid dynamics (CFD) to simulate reactor flow fields and on scale-down reactors to experimentally investigate selected scenarios (Kelly, 2008; Takors, 2012). Various designs of scale-down reactors exist and have been extensively reviewed elsewhere (Delvigne, Takors, et al., 2017; Delvigne, Destain, et al., 2006; Neubauer and Junne, 2010). One of the commonly used scale-down reactors follows a multi-compartment approach: A primary stirred tank

reactor (STR) is coupled to a secondary plug flow reactor (PFR). The STR is operated as a well-mixed compartment under standard limited growth conditions and the PFR simulates a feeding, starvation or anaerobic zone providing the stimulus to be investigated (Lara, Galindo, et al., 2006).

Many studies have focused on experimentally simulating the zone close to the feeding point which is usually characterized by substrate excess and potentially oxygen limitation (Enfors et al., 2001; Junne et al., 2011; Lara, Taymaz-Nikerel, et al., 2009). For a variety of hosts, common observations in this scenario include the formation of small organic acids and solvents as overflow metabolites or as anaerobic fermentation products (George et al., 1993; Neubauer, Häggström, et al., 1995). Ultimately, byproduct formation may lead to process performance loss even if reuptake of byproducts occurs in the well-mixed limited growth zone (Enfors et al., 2001).

Occasionally, starvation zones have attracted attention as well (Neubauer, Häggström, et al., 1995; Neubauer, Åhman, et al., 1995). From CFD simulation and measured data it is known that distant from the feeding point or close to the reactor walls poorly mixed zones with very low nutrient concentrations exist. An early scale-down study with *E. coli* employing oscillatory feeding protocols revealed the involvement of the stringent response in the cellular reaction to transient glucose starvation (Neubauer, Åhman, et al., 1995).

The stringent response is a global regulatory program usually preparing *E. coli* for entry into the stationary phase (Magnusson et al., 2005; Gaca et al., 2015; Hauryliuk et al., 2015). Its hallmark is the synthesis of the alarmone (p)ppGpp on short time-scales by the ribosome-associated protein RelA or on longer time-scales by the bifunctional enzyme SpoT (Atherly, 1979; Gallant et al., 1970; Murray and Bremer, 1996). ppGpp acts primarily as a transcription factor by binding to RNA polymerase and modulating its affinity to transcription initiation sites and alternative sigma factors. Additionally, ppGpp directly modulates the activity of certain proteins (Dalebroux et al., 2012; Kanjee, Gutsche, et al., 2011).

The fast and reversible initiation of the stringent response to oscillatory substrate supply was later confirmed by measurements of ppGpp in continuous glucose chemostat cultivations in a two-compartment stirred tank-plug flow reactor (STR-PFR) setup (Löffler, Simen, Jäger, et al., 2016b). The feeding point was placed in the STR creating a starvation zone in the PFR, which allowed to resolve the timescale of cellular response. Moreover, it was shown that extensive transcriptional responses take place as cells move transiently through a nutrient poor zone. From theoretical calculations of ATP costs Löffler *et al.* estimated that an increase in maintenance energy demand of more than 30% was caused

by the repeated exposure of cells to the nutrient gradient offering a new explanation for performance losses in large-scale bioprocesses (Löffler, Simen, Jäger, et al., 2016b). Analogous experiments with ammonium as the limiting nutrient revealed similar, yet less pronounced, regulation patterns affirming the importance of the stringent response for global regulation in *E. coli* in a scenario of oscillating starvation stimuli (Simen et al., 2017a). Fed-batch processes limited by ammonium or other nitrogen sources are interesting fermentation scenarios for the production of small molecules which mainly consist of carbon such as fatty alcohols (Chubukov et al., 2017). Nitrogen limitation is commonly used to enhance the accumulation of cellular carbon storage products such as polyhydroxyalkanoates used for bioplastic synthesis (Oliveira-Filho et al., 2019; Wen et al., 2010), including *E. coli* as a potential host (Wang, Yu, et al., 2009). As nitrogen forms a relatively large part of cells, nitrogen limitation can be easily explored during process development. During scale-up, such processes will likely suffer from similar issues as carbon-limited processes (Simen et al., 2017a).

Recently, the strains *E. coli* SR and *E. coli* HGT with modulated stringent response were constructed in our laboratory (Michalowski et al., 2017). The strains lack *relA* which is primarily responsible for rapid ppGpp synthesis upon nutrient depletion and carry modifications in the bifunctional enzyme SpoT. It was shown that they do not react to the exhaustion of ammonium supply by ppGpp synthesis (Michalowski et al., 2017). Strain *E. coli* SR displays no negative phenotypic differences in batch cultivations compared to its parent strain *E. coli* K-12 MG1655. However, under conditions of ammonium limitation, *E. coli* SR was found to have an elevated specific glucose consumption rate which is beneficial for two-stage processes involving product formation in the nitrogen limited phase (Jarmander et al., 2015; Perez-Zabaleta et al., 2016).

The combination of properties displayed by *E. coli SR* indicates that this strain can potentially be developed as a platform strain for robust scale-up from lab to production. In this work, we compared the phenotypic and transcriptional responses of *E. coli* SR and its parent strain *E. coli* MG1655 in a two-compartment scale-down reactor. We focused our investigation on the regulatory differences between these strains in the response to repeated short-term stimuli. The primary stirred tank reactor was operated as an ammonium-limited chemostat while a plug flow reactor simulated a nitrogen starvation zone.

5.3 Materials and Methods

5.3.1 Bacterial Strains and Media

Strains *E. coli* MG1655 or *E. coli* SR were used in all experiments (table 5.1).

2xYT agar plates were prepared by autoclaving 16 gL^{-1} tryptone, 10 gL^{-1} yeast extract, 5 gL^{-1} NaCl and 18 gL^{-1} agar-agar dissolved in demineralized water. Minimal media for precultures consisted of 4 gL^{-1} glucose, 0.96 gL^{-1} $NaH_2PO_4.2H_2O$, 3.51 gL^{-1} K_2HPO_4, 2.4 gL^{-1} $(NH_4)_2SO_4$, 0.01 gL^{-1} thiamine hydrochloride and 0.2% (V/V) trace elements stock solution. Minimal media for batch cultivation in the bioreactor consisted of 19 gL^{-1} glucose, 1.50 gL^{-1} $NaH_2PO_4.2H_2O$, 3.9 g/l K_2HPO_4, 5.7 gL^{-1} $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. 200 µl of antifoaming agent Struktol J647 (Schill + Seilacher, Hamburg, Germany) was added to the batch medium prior to inoculation. Minimal media for continuous chemostat cultivation in the bioreactor consisted of 11.4 gL^{-1} glucose, 1 gL^{-1} $NaH_2PO_4.2H_2O$, 2.6 gL^{-1} K_2HPO_4, 2.28 gL^{-1} $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. Throughout the chemostat phase 50 μlh^{-1} of antifoaming agent Struktol J647 were added continuously to the fermentation medium. The composition of trace element stock solution was 4.175 $FeCl_3.6H_2O$, 0.045 gL^{-1} $ZnSO_4.7H_2O$, 0.025 gL^{-1} $MnSO_4.H_2O$, 0.4 gL^{-1} $CuSO_4.5H_2O$, 0.045 $CoCl_2.6H_2O$, 2.2 gL^{-1} $CaCl_2.2H_2O$, 50 gL^{-1} $MgSO_4.7H_2O$ and 55 g/l sodium citrate dihydrate. Stock solutions of salts, trace elements and glucose were autoclaved separately, and stock solutions thiamine hydrochloride were filter sterilized and stored at 4 °C. All compounds were combined just before the experiments to prevent possible aging of media.

TABLE 5.1: Bacterial Strains used in this study.

Strain	Genotype/Strain Information	Reference
E. coli K-12 MG1655	F^-, λ^-, *ilvG*$^-$, *rfb*-50, *rph*-1 ('wild type' strain, abbrev. WT)	Michalowski et al., 2017
E. coli SR MG1655	Δ*relA*, *spoT*[R290E;K292D]	Michalowski et al., 2017

5.3.2 Bioreactor Setup

Cultivations were carried out in a two-compartment scale-down reactor. The primary reactor was a stirred tank reactor (STR), and a plug flow reactor (PFR) was used as the secondary compartment mimicking a starvation zone. The plug flow reactor was connected to the stirred tank reactor after establishment and sampling of a steady state in

the chemostat phase. The basic technical setup has been characterized previously (Löffler, Simen, Jäger, et al., 2016b; Simen et al., 2017a). Minor modifications to the original setup have been made and are described elsewhere (Ankenbauer et al., 2020).

The primary reactor was a 3 l bioreactor (Bioengineering, Wald, Switzerland) equipped with flow baffles and two six-blade Rushton type impellers operated at 1000 rpm. A constant aeration rate of 2.0 standard liters of ambient pressurized air per minute was employed and the system operated at a total pressure of 1.5 bar. Temperature was monitored by a platinum resistance thermometer and regulated by electrical heating or water cooling. Temperature was set to 28 - 30 °C for the batch phase and to 37 °C for the continuous chemostat phase. The reactor was equipped with a pH sensor (Mettler Toledo, Columbus, USA) to control pH and a pO2 sensor for monitoring dissolved oxygen tension (PreSens, Regensburg, Germany). During all fermentation stages pH was set to 7.0 and regulated by automated addition of 3 M NaOH or 2.5 M H_3PO_4. Dissolved oxygen tension was not regulated but maintained values above 70% saturation to 1.5 bar ambient air throughout the entire cultivation. In the exhaust gas stream, the concentration of oxygen and carbon dioxide was measured by gas sensors (BlueSens, Herten, Germany). During the chemostat phase the feed was constantly added to the reactor by a peristaltic pump (Watson-Marlow, Falmouth, United Kingdom). The feed flow was monitored by a balance recording the weight of the stirred feed barrel and manually adjusted if necessary. The harvesting pump operated as a slave pump set to maintain a constant weight of the bioreactor. For this purpose, the stirred tank reactor was installed on a balance as well.

The secondary compartment was a plug-flow reactor with an inner tube diameter of 20 mm and a total volume of approximately 380 ml. Five ports along the primary axis were used to take samples throughout the cultivation. Oxygen saturation in the PFR was monitored close to ports P1, P2 and P5 and additional aeration of 0.15 standard liters per minute was provided next to port P1 to ensure levels above 30 % saturation to ambient air conditions throughout the entire PFR passage. Temperature in the PFR was maintained at 36 - 37 °C by water heating and isolation material. A diaphragm metering pump (Sigma/1, ProMinent, Heidelberg, Germany) was used to transfer biosuspension from the stirred tank reactor to the plug flow reactor after connection of the two reactors.

5.3.3 Preculture, Batch Cultivation and Continuous Cultivation

A small amount of glycerol stock seed culture was spread onto 2xYT agar plates and incubated at 37 °C for 24 h. A single colony was picked to inoculate 500 ml baffled shaking flasks with 50 ml of preculture minimal media. Flasks were then incubated at 37 °C on an

orbital shaker set to 150 rpm for 16 hours. In the next morning 500 µl of biosuspension were transferred to 1000 ml baffled shaking flasks containing 100 ml preculture minimal media and incubated at 37 °C on an orbital shaker set to 150 rpm for 8 hours. 50 ml of this culture were used to inoculate the bioreactor. Total volume in the bioreactor was 1.6 l after inoculation. Batch fermentation in the bioreactor ensued at 28-30 °C overnight. In the next morning feed and harvest trains were connected and a constant feed/harvest rate at 5.33 $mLmin^{-1}$ corresponding to a dilution rate of 0.2 h^{-1} established. After 25 h (five volumetric residence times) of STR cultivation a reference sample was taken. The plug-flow reactor was then connected to the primary reactor via a diaphragm metering pump effectively circulating about one-quarter of the total fermentation broth from the STR through the PFR and back into the STR. In the following 28 h samples were taken at predefined time points from the STR and the five PFR ports. After 28 h of STR-PFR cultivation the fermentation was aborted, and the final broth volume measured. This value was used for all volumetric calculations during data analysis.

5.3.4 Determination of Optical Density and Biomass

In preliminary experiments with identical setup correlation factors of optical density and biomass as cell dry weight (CDW) were determined for *E. coli* MG1655 and *E. coli* SR (supplementary information A, Table C.1). The resulting correlation factors for converting OD_{600nm} values to gL^{-1} cell dry weight were 0.324 for *E. coli* MG1655 and 0.321 for *E. coli* SR. In the main cultivations optical density was measured from appropriately diluted broth on a spectrophotometer at 600 nm and converted into biomass concentration.

5.3.5 Determination of Acetic acid, Ammonium and Glucose Concentrations

5 ml of biosuspension was directly sampled into a syringe connected to a single-use 0.45 µm sterile filter and immediately sterile filtered. The clear supernatant was flash frozen in liquid nitrogen and stored at -70 °C until analysis. Glucose concentration was determined by D-Glucose UV-Test Kit (R-Biopharm, Darmstadt, Germany) and acetic acid concentration by Acetic acid UV-Test Kit (R-Biopharm, Darmstadt, Germany). Ammonium concentration was determined by Ammonium cuvette test LCK 304 (Hach Lange, Düsseldorf, Germany). At the end of the cultivation feed samples were taken and processed identically.

5.3.6 Analysis of Total Carbon, Inorganic Carbon and Biomass Composition

For total carbon and inorganic carbon analysis 0.5 ml biosuspension sample were mixed with 50 µl of 5 M KOH to prevent loss of dissolved carbonate. The suspension was then diluted 1:20 with demineralized water and stored at 4 °C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) to yield the total concentration of carbon and inorganic carbon in the fermenter effluent stream. At the end of the cultivation feed samples were taken and processed identically.

To determine biomass composition 1.0 ml of biosuspension was centrifuged at 4 °C and 14000 rpm (20817 g) for 3 min. The supernatant was discarded, the pellet resuspended in 1.0 ml of freshly prepared 0.9 % NaCl solution and centrifuged again. The pellet was resuspended in 5 ml 0.9 % NaCl, flash frozen in liquid nitrogen and stored at -70 °C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) and the carbon content of the biomass calculated from these values.

5.3.7 Measurement of ppGpp

2 ml of biosuspension was sampled directly into 0.5 ml precooled (< -20°C) quenching solution and incubated at 6 °C on a shaker for 15 min. Quenching solution consisted of 80 µM EDTA dissolved in 35 % (V/V) perchloric acid. 500 µl 1M K_2HPO4 was added and the sample briefly vortexed. 550 µl 5 M KOH was added and the sample vortexed again. To remove precipitating potassium perchlorate samples were then centrifuged at 4 °C and 7830 rpm (7197 g) for 5 min. 1.5 ml of supernatant was carefully transferred to new tubes, flash frozen in liquid nitrogen and stored at -70 °C. Prior to analysis samples were thawed and their pH adjusted to 6.95 - 7.05 with 5 M KOH or 35 % (V/V) perchloric acid. Samples were centrifuged again to remove all potassium perchlorate precipitate. HPLC analysis was carried out as described previously (Löffler, Simen, Jäger, et al., 2016b). If necessary, quantification was conducted by ppGpp standard addition (TriLink, California, USA). Samples from one time-point were analyzed directly in sequence and the data normalized to the sample drawn from the STR to eliminate differences caused by column aging.

5.3.8 Transcriptome Analysis

0.5 ml broth was sampled from the bioreactor and directly flash-frozen in liquid nitrogen. Frozen broth was then stored at -70 °C until the day of RNA isolation. Total RNA was isolated using RNeasy Mini Kit (Qiagen, Germany) according to the manufacturer's

instructions. Isolated RNA was DNAse treated and shipped to commercial sequencing partner GENEWIZ on dry ice. Samples were treated for rRNA depletion, sequencing libraries prepared and Illumina HiSeq 2x150 bp sequencing performed. Raw FASTQ files were obtained for bioinformatic analysis. Trimmomatic v. 0.32 (Bolger et al., 2014) was used to remove adapters and low-quality reads (<Q20) checked by fastqc reports. Genes were aligned to the NCBI *E. coli* K-12 MG1655 reference genome (RefSeq: NC_000913.3) using the RNA-sequencing aligner Bowtie2 v. 2.3.2.2 (Langmead et al., 2012). On average the mapping of the reads covered 96.2 %. Aligned reads were counted for each gene based on the corresponding annotation available from the NCBI database for the chosen reference sequence applying HTseq-count v. 0.6.1 in the union mode (Anders, Pyl, et al., 2015). On average 86.4 % of the sequenced reads could be assigned uniquely to annotated features. Sequencing depth was around 27 million reads per sample on average with a mean quality phred score of 37.63.

Differential gene expression analysis was performed with the R-package DeSeq2 v. 1.26.0 (Love et al., 2014) available from Bioconductor (Gentleman et al., 2004). Prior to statistical analysis, all residual non-protein encoding RNA molecules (tRNA, rRNA and sRNA) were removed from the HTseq-derived raw count data and a non-specific filter was applied to remove low coverage genes with fewer than two counts per million (54 reads on average). All filtering steps caused deviations from the raw data of less than 6 %. Samples were grouped by replicates and an experimental design was chosen that used sample time and location (STR or PFR port 5) as a combined environmental factor. To normalize read counts for the comparison of sequencing depth and RNA composition, DESeq2 uses the median of ratios method to derive a scaling factor. Dividing the original read counts by the scaling factor generated normalized count values. No outliers were observed in the two biological replicates using pearson correlation. Resulting p-values were adjusted for multiple testing according to control the false discovery rate (FDR) (Benjamini et al., 1995). Genes were identified as significantly differentially expressed by applying FDR adjusted p-values < 0.01 and a $\log_2$ fold change $\geq |1|$.

A principal component analysis was used to display the sample to sample distances calculated within the DESeq2 package (negative binomial distribution model). Principal component analysis was performed using plotPCA.san available on Github (https://gist.github.com/sansense/3399064897f1252d31b23ea5178c033c).

Gene set enrichment and overrepresentation analysis of up- and downregulated genes were performed using the Bioconductors's R-package GAGE v. 2.36.0 (Luo, Friedman, et

al., 2009). GAGE tests whether the mean fold-change of a gene subset is significantly different from the background using a two-tailed t-test. Genes were selected as significantly different with an FDR adjusted p-value < 0.01 (Benjamini et al., 1995). Functional annotation were derived from the Cluster of Orthologous Groups (COG) database (Tatusov et al., 2003), the experimental sigma factor-gene interaction dataset from RegulonDB v. 10.6.3 (Santos-Zavaleta et al., 2019) and the Gene Onthology (GO) Groups database with the function go.gsets from GAGE (Luo, Friedman, et al., 2009). Furthermore, Venn diagrams were used to identify significant genes shared by both strains and differences in gene expression regulation (Chen and Boutros, 2011). The RNA sequencing data derived from periodic ammonia starvation experiments have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GSE158198 (Edgar et al., 2002). Raw counts and processed data can be found in the Supplementary information. Data analysis was performed using the free statistical computing environment R v. 3.6.2.

5.4 Results

5.4.1 Continuous cultivation with periodic nutrient depletion

We cultivated *E. coli* SR and *E. coli* MG1655 in two independent continuous fermentations each in a previously described scale-down reactor consisting of a primary stirred tank reactor (STR) and a secondary plug-flow reactor (PFR), schematically shown in Fig. 5.1 (Löffler, Simen, Jäger, et al., 2016b; Simen et al., 2017a; Ankenbauer et al., 2020). *E. coli* SR is a strain with modulated stringent response that was engineered to alleviate the induction of the stringent response and the general stress response upon nutrient depletion (Michalowski et al., 2017). The chemostat was operated at a dilution rate of D = 0.2 h^{-1} and ammonium was chosen as the limiting nutrient. After establishment of a steady state in the STR alone, a reference sample (S0, t = 0 h) was taken and the PFR connected. Periodic passage from the STR (average residence time $\overline{\tau_{STR}} = 6.2$ min) through the PFR (average residence time $\overline{\tau_{PFR}} = 2.6$ min) then created a repeated short nitrogen starvation stimulus. The average residence times represent worst-case scenarios that are still consistent with mixing studies (Noorman, 2011; Vrabel et al., 2000) and the volume ratio STR to PFR was approximately 3:1 to represent existing simulation results (Haringa, Deshmukh, et al., 2017; Lapin, Schmid, et al., 2006). The long-term response of cells was investigated from additional samples taken from the STR shortly after connection of the PFR (S5, t = 5 min) and after establishment of a new stady-state (S28, t = 28 h) in the two-compartment

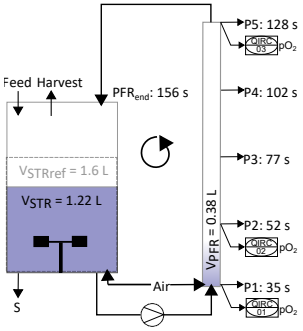


FIGURE 5.1: Experimental design of the two-compartment system. The fermenter consists of a stirred tank reactor (STR) as the primary cultivation vessel and a plug-flow reactor (PFR) connected by an active pump. The ammonium-limited chemostat was operated at a dilution rate of $D = 0.2\,h^{-1}$ with the feeding point placed in the STR. The STR served as a limitation zone and the PFR formed a starvation zone. The setup was designed to resolve different timescales of cellular response. Oxygen saturation was measured by 3 oxygen probes and recorded by the process control system (01, 02, 03). V_{STRref}: Reference Volume without connection of PFR (constant volume)

cultivation. The short-term response of cells to the PFR stimulus was monitored by sampling from five ports along the primary axis of the PFR at identical timepoints. Transcript samples for the PFR were taken from port 5 (P5 5 min and P5 28 h).

Basic growth and fermentation data confirmed earlier results that there are no detrimental differences in fundamental physiological parameters (Table 5.2) between *E. coli* MG1655 and *E. coli* SR under nitrogen-limited conditions (Michalowski et al., 2017). There were no statistically significant differences in any parameter (two-tailed t-test, $p > 0.1$). Both strains reached practically identical biomass yields on ammonium and depleted ammonium to equally low levels regardless of process time and PFR action (Fig. 5.2). The most noteworthy difference between *E. coli* MG1655 and *E. coli* SR was a reduced concentration of excess glucose in the fermentation broth of *E. coli* SR. Consequently, we calculated a lower biomass yield on glucose for *E. coli* SR (Table 5.2). Under conditions of long-term nitrogen starvation in batch fermentations *E. coli* SR had previously displayed a relaxation in glucose and nitrogen uptake coupling and we thus suspected an increased specific glucose uptake rate (Michalowski et al., 2017). The calculated specific glucose uptake rate was higher for *E. coli* SR, but the difference was not statistically significant in our experiments (two-tailed t-test, p-value > 0.1). Data from the fermentation broth supernatant showed that both strains converted comparable amounts of substrate into

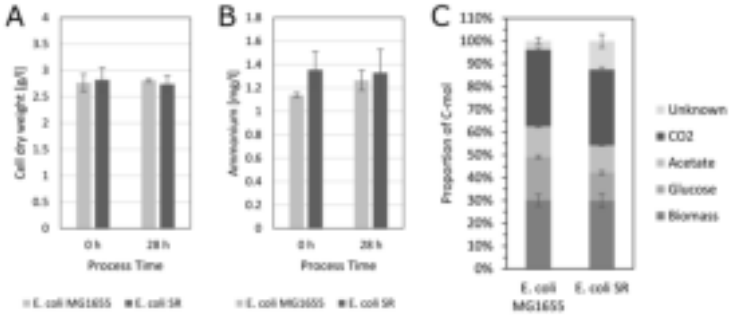


FIGURE 5.2: Physiological measurements. A: Cell dry weight. Concentration of cell dry weight after at least 25 h chemostat process before connecting the plug-flow reactor (0 h) and after 28 h of chemostat process with connected PFR (28 h). B: Ammonium. Concentration of residual ammonium in the supernatant. C: Carbon Balance. Columns show efflux fractions of total C-mol based on carbon influx. The final fraction represents undetermined dissolved organic substances in the fermentation broth, as measured by the difference of all efflux carbon detected by exhaust gas or total carbon analysis and the sum of the individually measured efflux components. Error bars indicate SEM (n = 2) of individual components (A, B and C).

acetate as the primary byproduct. Carbon balancing revealed an increased fraction of unknown substances among the fermentation products of *E. coli* SR which were identified as dissolved organic substances in the fermentation supernatant by total dissolved carbon analysis. The elevated glucose uptake rate of *E. coli* SR likely leads to higher byproduct formation of typical overflow metabolites such as lactate, pyruvate, formate and the regulator 2-oxoglutarate, all of which are known to accumulate under nitrogen-limited conditions with glucose excess (Hua et al., 2004). Apart from the primary byproduct acetate, individual small carbon byproducts were not measured as the overall total carbon efflux/influx balancing was in good agreement for both strains. Carbon recovery was 101 $\pm$ 2% for *E. coli* MG1655 and 102 $\pm$ 1% for *E. coli* SR indicating that in sum all relevant substances were detected.

In general, process time and the periodic PFR stimulus hardly affected global process parameters which is in accordance with former observations made in this reactor setup for nitrogen limitation and K-12 strains (Simen et al., 2017a). In sharp contrast, we found substantial regulatory differences between the two strains both in the short-term and in the long-term transcriptional responses to the periodic starvation stimulus.

TABLE 5.2: Physiological measurements.

Parameter	Unit	*E. coli* MG1655	*E. coli* SR
Y_{XN}	$\left[\frac{g_{CDW}}{g_{NH_4^+}}\right]$	4.63 ± 0.12^a	4.62 ± 0.27
Y_{XS}	$\left[\frac{g_{CDW}}{g_{Glucose}}\right]$	0.32 ± 0.01	0.28 ± 0.01
$c_{Glucose,STR}$	$\left[\frac{g_{Glucose}}{l}\right]$	2.07 ± 0.25	1.49 ± 0.06
$c_{Acetate,STR}$	$\left[\frac{g_{Acetate}}{l}\right]$	1.39 ± 0.11	1.29 ± 0.14
$q_{NH_4^+}$	$\left[\frac{g_{NH_4^+}}{g_{CDW} * h}\right]$	0.04 ± 0.01	0.05 ± 0.01
q_S	$\left[\frac{g_{Glucose}}{g_{CDW} * h}\right]$	0.63 ± 0.05	0.77 ± 0.14
q_{Ac}	$\left[\frac{g_{Acetate}}{g_{CDW} * h}\right]$	0.10 ± 0.01	0.10 ± 0.01
q_{CO_2}	$\left[\frac{mmol_{CO_2}}{g_{CDW} * h}\right]$	8.73 ± 1.06	9.98 ± 2.23
q_{O_2}	$\left[\frac{mmol_{O_2}}{g_{CDW} * h}\right]$	9.28 ± 0.47	10.9 ± 2.02
RQ	$\left[\frac{mol_{CO_2}}{mol_{O_2}}\right]$	0.95 ± 0.16	0.91 ± 0.04
q_{ATP}	$\left[\frac{mmol_{ATP}}{g_{CDW} * h}\right]$	29.23 ± 0.62^b	34.73 ± 6.39
D	$[h^{-1}]$	0.20 ± 0.01	0.21 ± 0.03

[a]Errors indicate SEM (n = 2). All rates were calculated from averaged values collected over the entire STR-PFR process time. [b]Estimated values assuming a P/O-Ratio of 1.2.

5.4.2 Transcriptomic analysis: Overview

RNA-seq-based transcriptomic data to examine potentially important genes for the ammonium stress response of *E. coli* WT and *E. coli* SR was analyzed. After filtering, 4037 predicted *E. coli* genes remained for further analysis (see supplementary data). The fast tactical transcriptional response to ammonia shortage was determined by comparing PFR port 5 samples to STR samples taken at the same process time points. Long-term responses were studied by comparing post-perturbation samples from the STR after 5 min (S5) and 28 h (S28) to the reference sample (S0). The statistical threshold for significance was set for adjusted p-value < 0.01 and $\log_2 FC > |1|$. 54 differentially expressed genes (DEGs) (UP: 14, DOWN: 40) formed the long-term response of *E. coli* MG1655. The short-term response was more pronounced comprising 837 DEGs (UP: 242, DOWN: 595). *E. coli* SR disclosed a similar number of 61 DEGs for the long term response (UP: 12, DOWN: 49), but substantially less DEGS as short term response (Total: 387, UP: 161, DOWN: 226) (Fig. 5.3). $Log_2 FC$ values range from -4.69 to 4.96 (WT) and -3.90 to 5.13 (SR). Fig. 5.3 depicts an overview of transcriptional dynamics outlining the halved response of *E. coli* SR 5 min after repeated nitrogen limited perturbation compared to WT.

Fig. 5.4 shows that the multi-transcript response of each strain could be well described

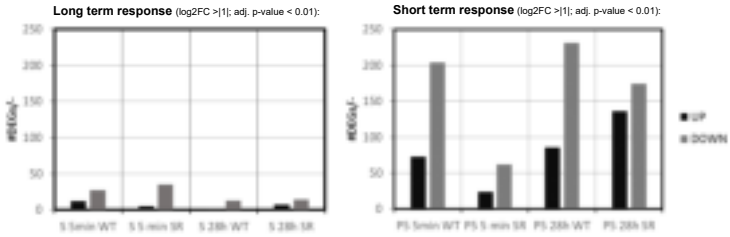


FIGURE 5.3: Number of UP (black) and DOWN (gray) regulated genes (DEGs). Long-term (left) and short-term (right) response to repeated nitrogen starvation for *E. coli* MG1655 (WT) and *E. coli* SR (SR) and given process times.

by 2-dimensional PCA covering 96% and 87% of total variance for *E. coli* WT and *E. coli* SR, respectively. Notably, biological duplicates were found in close proximity. PC1 accounts for the sample port location, PC2 for the time course. Unique and clearly distinguishable differences between STR and PFR transcript patterns were observed already after 5 min of repeated nitrogen starvation for both strains (Figure 5.4, C.1). In particular, principal component 1 (PC1) disclosed major differences between the samples of each strain accounting for 88% and 67% regarding *E. coli* WT and *E. coli* SR, respectively. The PCA finding is in agreement with the reduced number of DEGs observed for *E. coli* SR. The impact of PC2 is more pronounced for *E. coli* SR although almost identical numbers of DEGs were found as long-term response in both strains. However, given the low impact of PC1 for *E. coli* SR, similar DEG values affect the relative principal component analysis stronger.

As long-term responses of both strains were similar (see Appendix: supplementary information B) and weaker than short-term responses (Fig. 5.4) further analysis focused on short-term transcript patterns. Notably, changes between long- and short-term responses of both strains were dominated by counteracting transcript dynamics resetting perturbations after PFR passages (MG1655: 5 min and 28 h). Observations are in line with similar findings (Chang et al., 2002). Additional differences were found in the upregulation of carbohydrate transport (SR: 5 min) and catabolic processes (SR: 28 h) (see Fig. C.5 and C.6).

5.4.3 Regulatory response to short-term ammonium limitation

Preceding investigations of *E. coli* K-12 strains in STR-PFR scale-down reactors revealed the rapid accumulation of the alarmone ppGpp upon entry into the nutrient limited zone under both glucose and ammonium limitation (Löffler, Simen, Jäger, et al., 2016b; Simen et

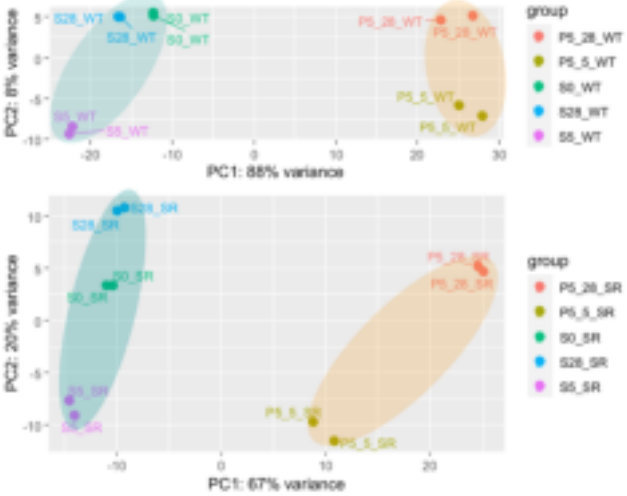


FIGURE 5.4: Principal component analysis of transcript data of *E. coli* MG1655 (WT) (**top**) and *E. coli* SR (**bottom**) obtained from STR (S) and PFR (port 5, P5) at three process time points (0 h, 5 min, and 28 h). Covered measurement variance of each principal component (PC) is indicated. Ellipses cluster samples of STR and PFR. PC1 accounts for 'sample port location', PC2 for 'process time'.

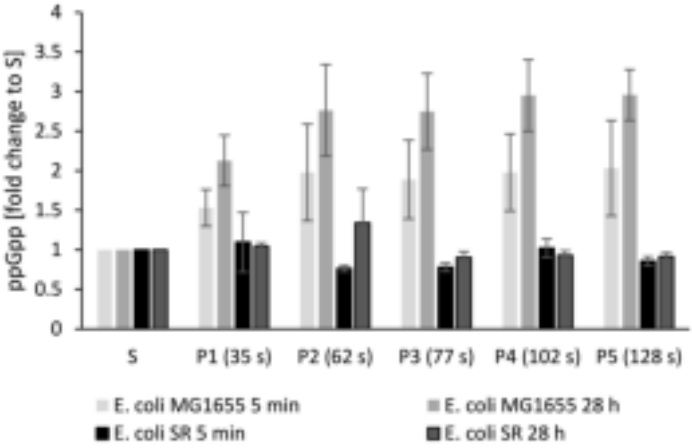


FIGURE 5.5: Alarmone accumulation along the PFR. Concentration of ppGpp measured from samples drawn along the plug flow reactor (P1 to P5) relative to the concentration measured in the stirred tank reactor (S, all values set to 1) for *E. coli* MG1655 (WT) and *E. coli* SR (SR). Error Bars represent SEM (n = 2).

al., 2017a). Concomitantly, an extensive transcriptional reprogramming of cells occurred. In standard batch fermentations *E. coli* SR in turn did not react to ammonium depletion by ppGpp synthesis (Michalowski et al., 2017). We therefore measured intracellular ppGpp levels from samples taken from the five ports of the PFR along its primary axis (Fig. 5.5). During the PFR passage *E. coli* MG1655 accumulated ppGpp to levels 2 - 3 fold higher than measured in the STR, displaying the same behavior as previously observed for the closely related K-12 strain *E. coli* W3110 (Simen et al., 2017a). In contrast, *E. coli* SR had no elevated levels of ppGpp at any point during the PFR passage regardless of process time. These results complement previous findings for the case of repeated short stimuli and confirm the strain's resilience to ammonium exhaustion.

Based on these encouraging findings, we focused our investigation on the short-term transcriptional response of both strains along the PFR axis. We compared data from samples drawn from port 5 of the PFR to samples drawn from the STR at identical process time points. Short-term changes revealed a significantly different response of *E. coli* SR compared to *E. coli* MG1655 not only in the amount of DEGs (Fig. 5.3), but also in the function of these genes (Fig. 5.4, 5.6). To elucidate patterns in the transcriptional responses, we searched for common DEGs, investigated the behaviour of gene clusters of orthologous

FIGURE 5.6: Venn diagrams representing partially overlapping sets of DEGs of *E. coli* MG1655 (WT) and *E. coli* SR. The number of significantly up- (UP) and downregulated (DOWN) genes in each set is indicated by numbers. Left: Short-term responses 5 min after PFR connection. Right: Short-term responses 28 h after PFR connection. Complete gene lists of the Venn diagrams are available in the supplementary data.

groups (COGs), and compared sigma factor (σ) activities. The gene expression patterns of each strain individually were assigned to 21 functional categories based on the COG database (Tatusov et al., 2003). In total 3532 of the 4037 genes (87.5%) could be annotated to COG. For each COG category, the resulting t-values are represented in a lollipop plot (Fig. 5.7). Significant changes were defined with a FDR-corrected p-value < 0.01. Furthermore, the activation and deactivation of sigma factors over time were investigated (Fig. 5.7). In this case, 3935 out of 4037 genes could be assigned to the sigma factor-gene interaction database from RegulonDB (Santos-Zavaleta et al., 2019).

After the first 5 min of PFR action *E. coli* MG1655 and *E. coli* SR exhibited substantially different transcriptional responses. The strains had only 64 DEGs in common, split equally between up- and down regulation (Fig. 5.6 left). Hence, these genes mirror the transcriptional response to short-term starvation irrespective of a functional stringent response, in which 14 out of the 32 common upregulated genes are associated with the Ntr-reponse (e.g. *glnK*, *amtB*, *glnAHPQ*, *rutA*). Downregulated genes consist of genes responsible for amino acid biosynthesis (e.g. *argCF*, *metABFINR*) and other cellular functions such as DNA cleavage, transporters, and oxidoreductases. The only oppositely regulated gene was *gua*C coding for the GMP reductase GuaC. Transcriptional control of the *gua*C promoter by the stringent response was proposed after its initial discovery and is clearly supported by our data (Andrews et al., 1988). Individual, strain-specific short-term regulation was observed for 398 (*E. coli* MG1655) and 32 (*E. coli* SR) specific DEGs after 5 min, clearly demonstrating the effect of the stringent response on the *E. coli* transcriptome.

Gene expression along the PFR after 28 h of PFR action differs strongly from the early response. 125 DEGs, mostly downregulated, are shared by both strains and the number of individually regulated genes is similar with 242 genes for *E. coli* MG1655 and 158

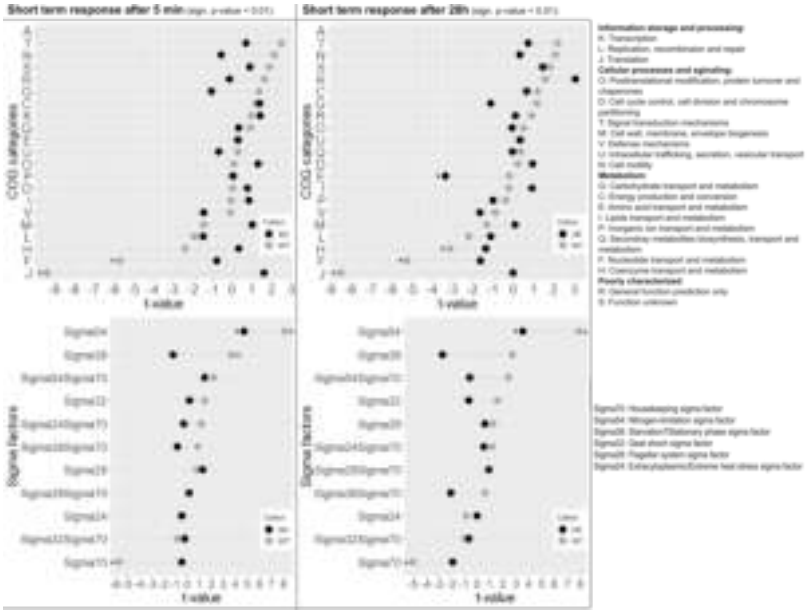


FIGURE 5.7: **Top**: Transcriptional patterns grouped into COG categories of *E. coli* MG1655 (WT) and *E. coli* SR (SR). Left: short-term patterns to the PFR stimulus 5 min after PFR connection. Right: short-term patterns to the PFR stimulus 28 h after PFR connection. **Bottom:** Sigma factor activities of *E. coli* MG1655 (WT, grey) and *E. coli* SR (SR, black). Left: Short-term response to the PFR stimulus 5 min after PFR connection. Right: Short-term response to the PFR stimulus 28 h after PFR connection. Significant categories are indicated with an asterix.

genes for *E. coli* SR (Fig. 5.6 right). Additionally, seven genes are oppositely regulated. Three of them (*tolQ, guaC, purM*) are upregulated in *E. coli* SR and downregulated in *E. coli* MG1655. These genes correspond to cell envelope integrity during cell division (Gerding et al., 2007), nucleotide metabolism (Kanjee, Ogata, et al., 2012) and purine *de novo* biosynthesis (Mueller et al., 1999). While purine *de novo* biosynthesis is actively inhibited by ppGpp via inhibition of GuaB, GTP synthesis solely originates from purine salvage pathways with *xdhA* significantly increased in *E. coli* MG1655 (Xi et al., 2000). The residual four oppositely regulated DEGs (*csiD, glnL, lhgO, yeaH*) predominantly play a role in the adaptation to nitrogen starvation and except for *glnL* are known to be induced by ppGpp. NtrB encoded by *glnL* is an essential part of the Ntr response cascade to nitrogen starvation and *yeaG* positively impacts *rpoS* transcription and translation under prolonged nitrogen starvation (Brown et al., 2014). Despite these differences in adaption to nitrogen limitation, we observed no alterations in the uptake or utilization of ammonium which indicates that the additional regulatory adaptions of *E. coli* MG1655 are irrelevant in the context of a bioprocess. Transcriptional patterns could be identified by functional enrichments of groups based on COG categories and sigma factor activities. COG groups J (Translation, ribosomal structure, and biogenesis) and F (nucleotide transport and metabolism) were significantly down regulated as part of the stringent response of *E. coli* MG1655 after both 5 min and 28 h (Fig. 5.7). For the 28 h sampling point group H (coenzyme transport and metabolism) was also significantly downregulated. As already indicated by the oppositely regulated genes (Fig. 5.7) σ54 mediated genes responsible for the activation of the Ntr stress response including *yeaG/H* via NtrBC were induced in *E. coli* MG1655, as well as the σ38 regulon as part of the general stress response (Brown et al., 2014; Figueira et al., 2015) (Fig. 5.7). Due to the limited amount of RNA-Polymerase (RNAP) core enzymes, σ70 competes with σ54, resulting in an antiproportional expression of their mediated genes (Jishage et al., 1996). In contrast, *E. coli* SR only increased the expression of genes regulated by σ54 after 5 min and no significant COG category was identified at this time-point. The absence of the stringent response in *E. coli* SR is clearly visible in an overall dampened regulatory response. The only significantly regulated group is E (amino acid transport and metabolism) after 28 h of PFR action, and the significantly downregulated genes in this group are predominantly ABC-transporters.

To unravel more detailed patterns in the transcriptional responses we assigned genes to the up-to-date gene ontology (GO) gene sets using GAGE (Luo, Friedman, et al., 2009). 3345 out of 4037 genes (83%) could be mapped to GO Terms. As shown in Fig. 5.3 the majority of significant DEGs for *E. coli* MG1655 were downregulated. This is mirrored by

the results of the identified top 20 GO categories which were uniformly down-regulated (Fig. 5.8). *E. coli* MG1655 predominantely downregulated genes related to ribosomal biosynthesis and translation after 5 min and 28 h as expected for a stringent phenotype (Fig. 5.8). These transcriptional changes are counteracted in the long-term response observed from the STR (Fig. C.3 - C.6) which indicates looping induction and repression of the genes. Patterns from *E. coli* SR were less pronounced and grouped differently. After 5 min we observed decreasing gene expression of ATP-demanding processes such as ABC transporters and ATPase complexes (Fig 5.8). After 28 h the PFR passage mainly induced an increased negative regulation of transcription and metabolic processes (Fig. 5.8). Care must be taken in the interpretation of this group though. General categories affecting transcription (GO:0006351, GO:0045892, GO:0097659, GO:1903507) or RNA processes (GO:0032774, GO:1902679, GO:0051253, GO:0051252) are represented as simultaneously negatively and positively regulated. Moreover, all negative regulators included in these terms, such as members of the CRP family, are also capable of positive regulation. Other negative regulation categories involve genes which actively inhibit translation and belong to SOS signals like DNA damage, prevention of cell division and programmed cell death (PCD). *E. coli* SR thereby focuses on σ38 regulated genes, as well as toxin and antitoxin systems (*mazEF* and *mqsRA*) possibly resulting in arrested growth and a dormant cell state or even PCD. As growth arrest is usually a primary outcome of the stringent response, which is absent in *E. coli* SR, we hypothesize that this pattern might provide an alternative way for *E. coli* SR to achieve cell cycle arrest.

In summary, the short-term response transcriptional patterns of *E. coli* MG1655 were extensive and dominated by the stringent response and the Ntr regulon. The major activated sigma factors were σ_{54} and σ_{38}. Overall, the transcription of ribosomal genes and other genes necessary for growth was inhibited, while genes involved in the transport and fixation of ammonia were induced. Our observations reflect well-known regulatory patterns exerted by *E. coli* K-12 when facing nitrogen starvation (Chang et al., 2002; Simen et al., 2017a; Traxler, Zacharia, et al., 2011; Traxler, Summers, et al., 2008; Wang and Levin, 2009). In contrast, the transcriptional short-term response of *E. coli* SR is dampened both in the number of DEGs and the patterns observed, especially shortly after connection of the PFR. The only significantly activated sigma factor is σ_{54} indicating a functional but attenuated Ntr response in the absence of ppGpp accumulation. Adaptation to ongoing starvation was possibly attempted via negative regulation of metabolic processes and SOS pathways.

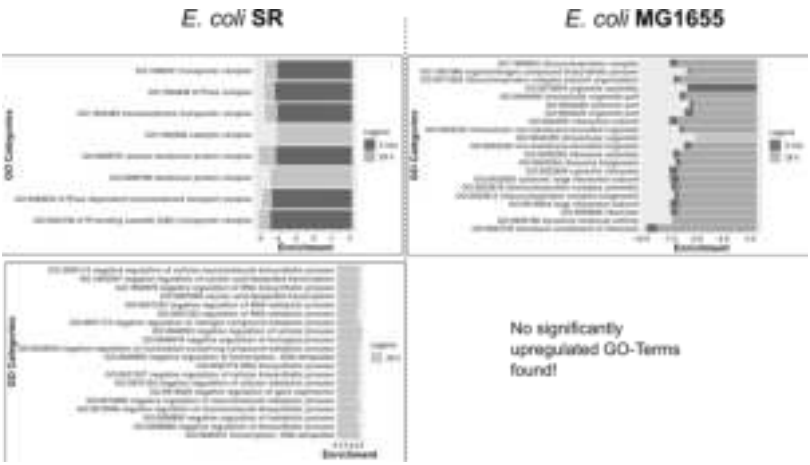


FIGURE 5.8: Significant GO categories after 5 min and 28 h of both *E. coli* SR (left) and *E. coli* MG1655 (right). Downregulated categories are arranged at the top and upregulated GO terms at the bottom. **5min:** Short-term response of *E. coli* SR (left) and *E. coli* MG1655 (right) after 5 min of PFR action. Only the Top 20 out of 102 significantly downregulated categories are shown. Neither strain had significantly upregulated categories for this time-point. **28h:** Short-term response of *E. coli* SR (left, light grey) and *E. coli* MG1655 (right, light grey) after 28 h of PFR action. For *E. coli* SR only the Top 20 out of 24 significantly upregulated categories are shown. For *E. coli* MG1655 only the Top 20 out of 95 significantly downregulated categories are shown. No significantly upregulated categories were found for this time-point.

5.5 Discussion

In the present study, we investigated the regulatory responses of the stringent response mutant strain *E. coli* SR when exposed to repeated short starvation stimuli in a scale-down reactor. The comparison with its wild-type parent *E. coli* MG1655 unraveled dampened regulatory patterns which are potentially beneficial for the application of *E. coli* SR in industrial large-scale reactors. The reduced regulatory patterns might be beneficial for heterologous protein expression as well as the production of small molecules as less interference with engineered metabolic pathways may occur and energy otherwise spent for adaptive responses is available for product formation.

An important finding of our study is that despite the regulatory differences *E. coli* SR displayed no dysfunctionalities in handling the shortage of ammonium. *E. coli* SR reached the same biomass yield on ammonium as *E. coli* MG1655 both with and without PFR action. Moreover, both strains depleted ammonium to comparable levels of about 1.2 mgL^{-1} or 67 $\mu molL^{-1}$, well in line with previously reported values for *E. coli* K-12 strains in nitrogen limited chemostats (Hua et al., 2004). The low remaining ammonium concentration indicates that uptake in both strains is mediated actively by AmtB with K_m = 0.8 mM (Williamson et al., 2020) and incorporation is accomplished by the GS-GOGAT System with GS K_m = 0.1 mM (Alibhai et al., 1994). This is supported by our transcriptional data which revealed that *amtB*, *gltB* and *gltD* were significanty enriched for both strains over all time-points. Transcripts of *glnA* were also always significantly enriched except for the time point 28 h of *E. coli* SR. Concomittantly, we identified transcriptional patterns typical for the σ_{54} and NtrBC mediated responses to nitrogen starvation (Reitzer, 2003). 13 out of 21 known NtrC-regulated operons (Brown et al., 2014) were induced at PFR port 5 in *E. coli* MG1655 at all time points (Table C.2). For *E. coli* SR, the Ntr response was slightly reduced, with 9 out of 21 operons induced (Table C.3) and lower overexpression of σ_{54} transcribed genes. These findings lead to the conclusion of an active, but diminished Ntr response of *E. coli* SR that still allowed fully functional ammonium assimilation. Additionally, the energy consumption as maintenance add-on for both strains was calculated according to Löffler, Simen, Jäger, et al. (2016b) assuming *de novo* synthesis of all upregulated DEGs over the whole process time (28 h). The resulting energy savings of *E. coli* SR due to weaker transcriptional response added up to around 46.5 %. In terms of microbial productivity, the reduced maintenance demand potentially increases the amount of available ATP for biomass-specific productivities and improves cell fitness.

In a previous study, a significantly elevated specific glucose consumption rate under

ammonium limitation was observed in *E. coli* SR (Michalowski et al., 2017). Similarly, we observed reduced excess glucose and the accompanying formation of dissolved byproducts in the fermentation supernatant. In *E. coli* K-12 strains, the consumption of glucose is usually tightly coupled to the availability of nitrogen on the level of metabolite control by the interaction of 2-oxoglutarate with PtsI (Doucette et al., 2011). The exact mechanism by which coupling of nitrogen and glucose uptake rates are relaxed in *E. coli* SR is not clear as the strain is isogenic to *E. coli* MG1655 except for the deletion of *relA* and the modifications in *spoT*. However, we found an increased transcription of *ptsI*, *ptsH* and *ptsG* in *E. coli* SR compared to *E. coli* MG1655 (Table C.4). Artifically increased expression of *ptsI* has been shown to increase specific glucose uptake rates in nitrogen limited conditions (Chubukov et al., 2017). We presume that the increased glucose uptake rate in *E. coli* SR might be caused by deregulated expression of *ptsI*, potentially connected to the absence of the stringent response by the action of CRP whose transcription is negatively regulated by ppGpp (Johansson et al., 2000). It remains to be clarified wether *E. coli* SR has altered cytoplasmic 2-oxoglutarate levels or the action of ppGpp influences the coupling of glucose consumption to nitrogen availability, potentially by the proposed mechanism. Increased specific glucose uptake rates in conjunction with higher respiratory activity have also been observed in *E. coli* MG1655 subjected to repeated glucose feast-famine cycles (Vasilakou et al., 2020). Future studies should thus examine how *E. coli* SR reacts to varying availability of glucose or other carbon sources.

In view of these differences in carbon metabolism, we hypothesized that biological energy availability might be unequal for *E. coli* MG1655 and *E. coli* SR. From oxygen and glucose uptake rates the specific ATP production rate q_{ATP} was estimated (Table 5.2). q_{ATP} greatly depends on the effective P/O ratio and current scientific consensus estimates realistic P/O ratios between 1.0 and 1.5 for *E. coli* (Noguchi et al., 2004; Szenk et al., 2017). For our estimations of q_{ATP} we assumed a conservative P/O ratio of 1.2 and 2 moles of ATP per mol glucose from glycolysis. The result indicates that *E. coli* SR might have an increased availability of ATP compared to its wild-type parent under the applied experimental conditions. Given that the respiratory capability and thus the ATP production capability of K-12 strains is not exhausted at a dilution rate of D = 0.2 h^{-1} it appears that the increased glycolytic flux to byproducts displayed by *E. coli* SR was also not a result of increased energy demand. Moreover, increased glucose uptake has been reported previously for *E. coli* SR under conditions of ammonia limitation despite high adenylate energy charge (Michalowski et al., 2017). Carbon and redox homeostasis at elevated glycolytic

flux would then be maintained by byproduct excretion and increased respiration, possibly involving the dissipation of surplus energy by uncoupling of the electron transport chain (Bekker et al., 2009).

Nitrogen limitation inducing the stringent response is a well-documented phenomenon in *E. coli*. Multiple previous studies predominantly observed heavily increased gene expression corresponding to amino acid transport and metabolism (Barker, Gaal, et al., 2001; Brown et al., 2014; Simen et al., 2017a; Traxler, Zacharia, et al., 2011; Traxler, Summers, et al., 2008; Durfee et al., 2008). Conversely, we observed almost equally distributed up- and downregulated genes for amino acid transport and metabolism (see supplementary data: Transcriptomics), which was only reported by few research groups (Chang et al., 2002; Traxler, Summers, et al., 2008). As a result, no overall significant statistical trend was detectable for this category (Fig. 5.7). We suggest that the individual operons do not solely respond to ppGpp, but rather depend on other signals and regulatory networks which were not found to be significantly expressed in this study such as the Lrp regulon. Additionally, caution is advised when comparing transcriptomic analyzes originating from different studies as they greatly depend on the transcriptional reference state and thus the details of the experimental design.

In general, the amount of DEGs of *E. coli* K-12 MG1655 was similar to the numbers found in the analogous study of Simen et al., 2017b who employed the closely related *E. coli* K-12 W3110 confirming the validity of our data. The amount of DEGs is also less than observed during the related study of glucose starvation by Löffler, Simen, Jäger, et al., 2016b which points towards significant potential of *E. coli* SR to preserve energy in glucose starvation conditions. An interesting difference to the former studies in this scale-down reactor setup is the absence of increased motility in the STR after PFR connection (Löffler, Simen, Jäger, et al., 2016b; Simen et al., 2017a). Our dataset contains no upregulated flagellar or sigma factor 28 mediated gene patterns from the STR at any time-point (Fig. 5.7). We first hypothesized that the cause might be genetic differences affecting motility which are well documented between MG1655 and W3110 and even between different MG1655 isolates (Barker, Prüß, et al., 2004; Hayashi et al., 2006). However, sequencing of our MG1655 isolate revealed the presence of the canonical IS-1 insertion upstream of *flhD* which confers motility and our MG1655 isolate displayed vivid spreading in motility agar (Supplementary information D, Fig. C.7). An alternative explanation could be derived by the interplay of quorum sensing and flagellar regulation through the action of autoinducer-2 (AI-2) and the motility quorum sensing regulator MqsR. While transcript levels of *luxS* (LuxS synthesizes Al-2) remain unchanged, the expression of *mqsR*

is significantly enriched at PFR port 5 and MqsR is known to induce the flagellar synthesis cascade (Barrios et al., 2006). However, cell dry weight (CDW) was always below 3 gL^{-1} in our experiments whereas Simen *et al.* worked with around 10 gL^{-1} CDW. Higher biomass should lead to increased AI-2 levels and may cause a preconditioned phenotype that rapidly initiates flagellar biosynthesis when encountering nutrient stress. Thus, rapid induction of motility genes might become more pronounced during high cell density processes in large-scale reactors and remains to be examined in further studies. Additionally, as introduced by Löffler, Simen, Jäger, et al. (2016b) during glucose fluctuation, genes of the category cell motility were identified as one of the most prominent energy consumers and might therefore be candidates for genome reduction (Löffler, Simen, Jäger, et al., 2016b).

Analysis of gene expression patterns (Fig. 5.7 and 5.8) revealed that both strains individually adapted to repeated nitrogen starvation. *E. coli* MG1655 adjusted by utilizing the ppGpp-mediated general stress response including activation of toxin/ antitoxin (TA) systems like *mqsRA* and *mazEF*. This strategy intends to arrest the cell cycle and form persister cells (Balaban et al., 2004). Persister cell formation is not yet fully understood and usually only involves a small fraction of cells (Chowdhury et al., 2016; Korch et al., 2015; Gerdes et al., 2012). Thus, it seems to be only of minor importance for industrial processes but some persister genes affect persister level due to altered growth rates rather than contributing to a mechanism of cell cycle arrest and might have a signgificant impact on bioprocess performance (Allison et al., 2011). Nonetheless two common dependencies affecting persister formation, ppGpp and TA systems, are known which is in line with our findings (Barrios et al., 2006; Chowdhury et al., 2016; Wang and Levin, 2009; Aizenman et al., 1996; Sun, Guo, et al., 2017). Persister formation benefits from increased ppGpp concentrations but is still possible at lower rates in the absence of ppGpp by proteins which simply reduce growth (Chowdhury et al., 2016). The nucleotide pyrophosphohydrolase MazG which is negatively regulated by the *mazEF* system is able to initiate cell cycle arrest and was significantly upregulated in *E. coli* SR after 28 h (Lee, Kim, et al., 2008). Additionally, *E. coli* SR initiated negative regulation of transcription, translation and cell division processes as part of the SOS response (Fig. 5.8). Most likely, the SOS pathways were activated due to ongoing DNA replication during starvation conditions which might ultimately result in DNA damage and inhibited cell division (Traxler, Summers, et al., 2008; Bi et al., 1993; Joseleau-Petit et al., 1999). As part of the SOS response and as a key gene involved in filamentation *sulA* was significantly upregulated in *E. coli* SR. SulA inhibits the initiation of cellular divison by repressing the assembly of FtsZ into the Z ring (Fonville et

al., 2010; Huisman et al., 1984). Simultaneously with the overexpression of *sulA*, *lexA* was significantly increased which acts as a major repressor of SOS signals. LexA regulates the response strength and is actively involved in the occurrence of persister cells in bacterial populations (Butala et al., 2011). These results indicate a coordinated and rather complex SOS response in *E. coli* SR to form persister cells which is not yet fully understood.

The natural regulation of *E. coli* has evolved towards optimality in its lifestyle as a gut bacterium and is not honed for the demands of a large-scale bioprocess. The absence of the stringent response and the conservation of the ability to grow efficiently in minimal medium suggest that *E. coli* SR has the potential to become a platform strain for applications in large-scale reactors. Our transcriptional analysis shows that the short-term response of *E. coli* SR to ammonium depletion is dampened but a functional Ntr/σ_{54} response remains. Regarding glucose-limited fermentations, we hypothesize that *E. coli* SR has significant potential to preserve energy in such conditions since the regulatory responses are usually even more pronounced and centered around the stringent response (Hardiman, Lemuth, Keller, et al., 2007; Löffler, Simen, Jäger, et al., 2016b). We therefore propose to confirm the suitability of *E. coli* SR for large-scale applications in multi-compartment scale-down reactors employing exemplary small-molecule production scenarios. These should include standard glucose-limited fed-batches as well as ammonium limited fed-batches with a prolonged nitrogen-limited production phase to exploit its elevated glucose consumption.

Chapter 6

Data-Driven *In-silico* Prediction of Regulation Heterogeneity and ATP Demands of *Escherichia coli* in Large-scale Bioreactors

Authors

Julia Zieringer (first author) - University of Stuttgart, Institute of Biochemical Engineering
Moritz Wild (second author) - University of Stuttgart, Institute of Biochemical Engineering
Ralf Takors (corresponding author) - University of Stuttgart, Institute of Biochemical Engineering

6.1 Abstract:

Escherichia coli exposed to industrial scale heterogeneous mixing conditions respond on external stress by initiating short-term metabolic and long-term strategic transcriptional programs. In native habitats, long-term strategies allow to survive severe stress but are of limited use in large bioreactors where micro environmental conditions may change right after said programs are started. Related on/off switching of genes causes additional ATP burden that may reduce the cellular capacity for producing the desired product. Here, we present an agent based data driven model linked to computational fluid dynamics finally

allowing to predict additional ATP needs of *E. coli* K12 W3110 exposed to realistic large-scale bioreactor conditions. The complex model describes transcriptional up- and down-regulation dynamics of about 600 genes starting from subminute range covering 28h. The data-based approach was extracted from comprehensive scale-down experiments. Simulating mixing and mass transfer conditions in a 54 m^3 stirred bioreactor, 120.000 *E. coli* cells were tracked while fluctuating between different zones of glucose availability. It was found that cellular ATP demands rise between 30 - 45% of growth decoupled maintenance needs which may limit the production of ATP-intensive product formation accordingly. Furthermore, spatial analysis of individual cell transcriptional patterns reveal very heterogeneous gene amplifications with hot spots of 50-80 % mRNA upregulation in the upper region of the bioreactor. The phenomenon reflects the time-delayed regulatory response of the cells that propagates through the stirred tank. After 4.2 h cells adapt to environmental changes but still have to bear additional 6 % ATP-demand.

6.2 Introduction:

To reduce the human CO_2 footprint in atmosphere sustainable bioprocesses replacing fossil resources by sugar may play a crucial role. Microbial production offers the potential to provide products for agricultural, biopharmaceutical and chemical markets (Delvigne, Takors, et al., 2017). As a prerequisite, such approaches need to be transferred successfully from laboratory to large-scale without loss of economic attraction, i.e. without reduction of the sensitive TRY values (titer, rates, yields) that served as constraints for economic evaluation. However, performance losses may occur comprising increased by-product formation, reduced substrate-to-product conversion, reduced productivities etc. (Lara, Galindo, et al., 2006). They mirror cellular responses to large scale heterogeneities that are induced by limited mass transfer and by insufficient mixing capacities (Noorman and Heijnen, 2017). Accordingly, research activities aimed to mimic large-scale conditions already in early stage lab tests. One of the first examples is given by Oosterhuis, Groesbeek, et al. (1983) who repeatedly exposed cells to oxygen saturated and limiting conditions in a setting of two linked, stirred bioreactors. Multiple tests with alternate experimental scale-up simulators followed (overviews provided in Garcia-Ochoa et al. (2009), Neubauer and Junne (2010), Neubauer and Junne (2016), Noorman (2011), Takors (2012), and Zieringer and Takors (2018)) mimicking not only fluctuations of dissolved oxygen (DO) levels, but also nutrient availability and pH variations. Today, such approaches received key consideration to design robust microbial processes (Noorman and Heijnen, 2017). Still, the

valid *a priori* prediction of large scale heterogeneities' impact on cellular performance is of crucial importance for developing novel bioprocesses. Even further, findings of large scale stress exposure may guide strain engineering to create particularly robust hosts. To reach this goal, Löffler, Simen, Jäger, et al. (2016b) applied the so called STR/PFR setup comprising a stirred tank reactor (STR) linked with a plug-flow reactor (PFR). Steady-state nutrient-limited, continuous cultivations were performed in STR before PFR was connected frequently exposing cells to glucose limiting conditions. Accordingly, cells repeatedly experienced temporal feast/famine conditions that were characterized by the residence time in the PFR. Comprehensive sampling in STR and PFR created a highly valuable dataset of short- and longterm metabolic and transcriptional responses on repeated starvation stimuli (Löffler, Simen, Jäger, et al., 2016b). The dataset revealed that *E. coli* not only reacts on extracellular stress by instantaneous metabolic shifts. Observations also revealed massive transcription of genes organized in operons and in fundamental regulons of strategic importance (Nieß et al., 2017). For instance, stringent response was repeatedly initiated by fast rising intracellular (p)ppGpp levels in PFR which were down-regulated in STR. Löffler, Simen, Jäger, et al. (2016b) reported additional rise of growth-decoupled maintenance of up to 50%. So far, these findings were not yet used to predict the response of *E. coli* exposed to large-scale heterogeneities. First, a data driven model is needed that describes the complex transcriptional response of *E. coli* to said stress conditions. Next, such a transcriptional model should be coupled with computational fluid dynamics (CFD) of a large scale bioreactor to identify zones of different nutrient availability and to predict the cellular response of cells passing through those zones (Zieringer and Takors, 2018). Our study exactly tackles this two step problem: Mixing heterogeneities and zones of different substrate availability of a 54m³ stirred bioreactor are predicted using CFD and assuming common operating conditions. The tracking of 120,000 *E. coli* cells finally yielded the prediction of additional ATP demands. Furthermore, spatially resolved transcriptional patterns of individual *E. coli* cells were predicted unraveling the population heterogeneity in the industrial-scale bioreactor.

6.3 Materials and methods

6.3.1 Experimental setup

A glucose gradient was simulated in a stirred tank reactor (STR) coupled to a plug flow reactor (PFR) as depicted in figure 6.1.

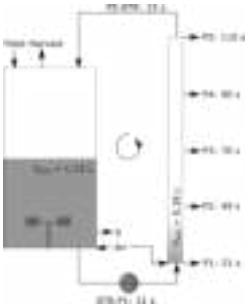


FIGURE 6.1: Scheme of two compartment system as used in the experimental setup (Löffler, Simen, Jäger, et al., 2016b). The two-compartment device consists of a stirred tank reactor (STR) connected to a plug-flow reactor (PFR). Derived from non-ideally mixed large-scale industrial fermenters (Lapin, Schmid, et al., 2006), the setup mimics periodic substrate availability experienced by cells in large-scale bioreactors. The well-mixed STR is operated in glucose limited continuous mode (Dilution rate D = 0.2 h^{-1}). As soon as cells enter the PFR compartment the residual substrate is consumed within seconds leading to starvation. The steady state prior to PFR onset at time zero was used as the reference state (S0). Samples were tak-en at eleven distinct time points over 28 h. The system is equipped with five PFR sample ports (P1-5) at defined residence times (s), as well as an STR sample port S. The total mean PFR residence time is PFR = 125 s.

The experimental setup consists of a stirred tank reactor operated in continuous mode (Dilution rate D = 0.2 h^{-1}) and connected to a plug flow reactor. Cells were grown under glucose limited conditions in the STR (mean residence time of the cells in STR: 6.2 min) and experience starvation in the PFR (mean residence time of the cells in PFR: 125 s). The cells are circulating through the reactor system for 28h process time which equals on average around 200 passages of the starvation zone for each cell. In this way, the setup permits the analysis of transcriptional response for ongoing starvation passages through the PFR. Thereby, the tactical response is monitored via the PFR sample ports (P1-P5), while the strategic changes were tracked via the STR sample port (S). The cultivation was performed as biological triplicates under identical experimental conditions. For transcriptomic analysis the samples were grouped by replicates and sample time and location (STR or PFR) was chosen as a combined experimental design. Significantly expressed genes were determined using the described design and a generalized linear model within the egdeR R-package (v.3.8.6) (Robinson and Oshlack, 2010). The detailed experimental implementation and RNA sequencing results used in this publication were published in the paper of Löffler, Simen, Jäger, et al. (2016b).

TABLE 6.1: Parameters and omitted genes of the SQBC algorithm. *Ntry*: number of trials per iteration; *rmax*: maximum radius as a proxy for correlation; *Min size*: minimum number of observations per clusters

Time point	25 min	2 h	28 h
ntry	1000	1000	1000
rmax	0.5	0.5	0.9
n_{min}	3	3	5
Genes omitted	0	1	6

6.3.2 RNA-sequencing data cluster analysis

RNA-sequencing data contain time courses of mRNA abundance of 3908 genes. Thereof, three measurement sets sampled after 25 min, 2 h and 28 h were chosen for further investigations. Significantly up- and down-regulated genes of samples P1-P5 in PFR passing the threshold of $\log_2$ fold change ($\log_2$FC) $> |0.58|$ and false discovery rate (FDR)-corrected p-value < 0.05 were identified. Cluster analysis was performed using the R-package *flexclust* v. 1.4-0 (Leisch, 2006) applying RStudio v. 1.2.1335 (RStudio, Inc.) to significantly reduce simulation efforts while including basic features of gene dynamics. The function *qtclust* (included in flexclust package) was used to perform stochastic quality based clustering (SQBC) and *k*-means-clustering. Parameters of SQBC were set as follows:

Ntry indicates the number of trials per iteration whereas *rmax* is the maximum radius as a proxy for correlation. N_{min} defines the minimum number of observations per clusters. Data points not clustered by the algorithm are omitted. The setting of parameters ensured maximum comparability and five as the maximum number of clusters. Cluster properties are listed in the appendix, tables D.3, D.4, D.5 and D.6 and resulting clusters are displayed in supporting information D.4 in the appendix. The *k*-means algorithm was initialized with the centroids of the SQCB method.

6.3.3 ATP calculation for single molecules

ATP requirements for the formation of amino acids and nucleotides were calculated using the results of Kaleta et al. (2013). The translational costs for protein formation and polymerization add up to 4 ATP per amino acid including activation of the amino acid (1 ATP to 1 AMP) and peptide bond formation at the ribosome (2 GTP) (Stouthamer, 1973). Since there is a net production of 0.1 ATP per amino acid (for detailed calculation see Löffler, Simen, Jäger, et al. (2016b), the overall cost of amino acid synthesis and polymerization was estimated as 4 ATPs consumed per residue. The absolute numbers of synthesized and degraded nucleotides (Nts) were estimated from experimental data. To recycle mono

phosphorylated nucleotides (NMPs) to triphosphorylated nucleotides (NTPs) costs of 2 ATP were assumed. Assuming a P/O-ratio of 1.49 (ATP formation via NADH oxidation in respiration), ATP requirements as shown in table 6.2 were assumed for the bases.

TABLE 6.2: ATP costs for *de novo* nucleotide synthesis, based on Löffler, Simen, Jäger, et al. (2016b)

Base	ATP	NADH	NADPH	overall ATP
Guanine	11	-3	1	8.53
Cytosine	13	-3	0	9
Adenine	9	0	1	6.53
Uracil	7	0	1	7
Average	10	-1.5	0.75	7.7

The growth-independent maintenance was used as 0.0027 mol $(g_{DW}h)^{-1}$ according to Taymaz-Nikerel, Borujeni, et al. (2010).

6.3.4 Calculation of mRNA abundance

Only additional ATP needs for transcription and translation were estimated considering the basic demands under non-perturbed conditions. Accordingly, total mRNA content was estimated following studies of Bremer et al. (2008) as 61.7 µg per 10^9 cells for growth rate of 0.2 h^{-1}. 20% of the total dry weight were assumed to be RNA (Neidhardt et al., 1990) including 5 % mRNA, a value which is in line with conclusions from Stouthamer (1973). This results in total mRNA content of 6.17g per 10^{16} mRNA per cell. The relative distribution of specific mRNAs is taken from the measured normalized counts as transcripts per million. The relative fraction multiplied with the total mRNA content gives the total mass of all mRNA encoded by a single gene. Dividing this number with the corresponding molecular weight yields the absolute number of molecules. Molecular weights of mRNAs (MW_{mRNA}) were calculated with equation 6.1 (Kibbe, 2007). Results are listed in Appendix Table D.1. As the phosphate groups of two nucleotides are bound together, an OH-group is cleaved. This leads to reduced molecular weights of nucleotides in the polymer chain. Accordingly, an additional term was added to account for the 5'-triphosphat cleavage (159 $gmol^{-1}$).

$$MW_{\mathrm{mRNA}} = n_{\mathrm{G}} * 345.2 + n_{\mathrm{C}} * 305.2 + n_{\mathrm{A}} * 329.2 + n_{\mathrm{U}} * 306.2 + 159 \tag{6.1}$$

$n_{\mathrm{nucleotide}}$ codes for the number of nucleotide monophosphates which is multiplied by their corresponding molecular weight.

6.3.5 The biological model

The model was implemented using Matlab v. 2019b considering the four processes: transcription, translation, mRNA degradation and protein degradation. The first three are implemented as agent-based approaches whereas the last considers protein degradation as a decomposition in continuum. The governing variable that controls the expression is the number of active RNA polymerases (RNAPs) per each cluster.

6.3.6 Estimating the number of active RNAP

We assumed that gene expression levels follow sigmoidal courses. Hence, an equilibrium between synthesis and degradation may be achieved. Produced nts are given by

$$Nt_{prod} = v_{\mathrm{RNAP}} \cdot \int_0^t \sigma_{\mathrm{RNAP}}(t)\,\mathrm{d}t \tag{6.2}$$

with $\sigma_{\mathrm{RNAP}}(t)$ coding for the number of active RNAPs at time t. The shape of the sigmoidal function is defined as

$$\sigma(t) = \frac{a}{1 + e^{\{b\cdot(t+c)\}}} + d \tag{6.3}$$

The parameters a, b, c and d were fitted to the nucleotide synthesis which was derived from experimental data by calculating the number of synthesized copies and by considering individual gene lengths. Consequently, steadily rising functions were obtained that allowed to estimate the number of active RNAPs per cluster. For the latter, a constant RNAP transcription velocity of 21 nucleotides per second was assumed (Chen, Shiroguchi, et al., 2015a).

6.3.7 Transcription

After initiation, continuous one-stranded movement of RNAP creates the mRNA transcripts measured. However, individual gene expression profiles were observed that could be grouped in clusters of similar transcription dynamic. Accordingly, only expression dynamics of representative, average genes per cluster are described in the model (Appendix table D.3-D.5). The minimum distance of 100 nts (Δx_{RNAP}) was considered between two subsequent RNAPs. Furthermore, all genes of one cluster were supposed to be randomly initiated with the functional

$$gene_{\mathrm{i,on|off}}(t) = \begin{cases} 1 & \text{if free RNAP is avaliable and } gene_{\mathrm{i}} \text{ is randomly chosen} \\ 0 & \text{else} \end{cases} \tag{6.4}$$

Gene transcription is modeled as one dimensional nucleotide extension with the relative movement of RNAP as

$$\frac{dx_{\mathrm{RNAP}}}{dt} = \begin{cases} v_{\mathrm{RNAP}} & \text{if initiated} \\ 0 & \text{else} \end{cases} \tag{6.5}$$

The constant transcriptional elongation rate, v_{RNAP}, is set to 21 nts^{-1} which equals the average value found in *E. coli* during starvation (Chen, Shiroguchi, et al., 2015a). The variable $x_{mathrmRNAP}$ indicates the relative position of RNAP on the DNA grid. The length of the resulting mRNA strand is equivalent as

$$x_{\mathrm{i,RNAP}} = L_{\mathrm{i,mRNA}} \tag{6.6}$$

When the last nucleotide is reached, the mRNA is released. All fragments of mRNA are summed to get total mRNA amounts. Whereas fractions of operons were found to be fully transcribed after initiation (Nieß et al., 2017), other scenarios coincided, too. For instance, only subsets of operons may be transcribed or even opposing transcription reads in a single operon occurred (Mao et al., 2015). Accordingly, we assumed that only 10% of the initiated operons are transcribed completely. In other words, 10% of experimentally observed initiated operons were anticipated to finish full operon transcription even outside of PFR. The majority (90 %) of other gene transcriptions was assumed to stop immediately after RNAPs have reached individual gene ends. For the sake of comparability only relative mRNA enrichments are depicted in figure 6.7 C referring to the mRNA level of individual cells after they have fluctuated through the bioreactor for 180 s. This time point was chosen to visualize the spatial distribution of already adapted cells and the ones which are still influenced by regime changes (figure 6.7).

6.3.8 Translation

The translational process is modeled by describing the movement of ribosomes (RIB) on the mRNA strand. The process is assumed to take place in co-transcriptional manner (Proshkin et al., 2010): After synthesized mRNA reached minimum length of 80 nts

(Δx_{RIB}) (Bremer et al., 2008), the first ribosome attaches to the free 5′-cap end. Further ribosomes may bind too, provided that minimum distance between two subsequent ribosomes and maximum number of translations per mRNA (max_{TL}) are fulfilled. Released ribosomes may be reused following the same scenario. The initiation trigger $TL_{\mathrm{i,j}}$for a translation j on mRNA i can be described as

$$TL_{\mathrm{i,j,on|off}}(t) = \begin{cases} 1 & \text{if position of } RIB_{i,j-1}\text{is} > 80 \text{ and j} < max_{\mathrm{TL}} \\ 0 & \text{else} \end{cases} \tag{6.7}$$

The movement of ribosomes is analogous to the movement of RNAP as

$$\frac{dx_{RIB}}{dt} = \begin{cases} v_{RIB} & \text{if initiated} \\ 0 & \text{else} \end{cases} \tag{6.8}$$

With the translational elongation rate v_{RIB}= v_{RNAP}(Nieß et al., 2017, Proshkin et al., 2010) and the number of maximum translations max_{TL} = 11 (Bremer et al., 2008). Because one cluster of genes revealed rapid degradation after 50 s in PFR at 28 h, only 1 translation per transcript was assumed for this group of genes. As soon as the ribosomes passes the last nucleotide, the protein is released and is assigned to the group of bulk proteins.

6.3.9 Degradation

Degradation of transcripts is initiated as soon as ribosomal protection of the 5′-cap vanishes which is a co-translational process. The velocity of the RNASE was adapted to the gene length $L_{mRNA,i}$, i.e. individual $v_{\mathrm{RNASE,i}}$ were estimated considering the experimentally observed mRNA median life time $t_{\mathrm{deg;med}}$ of 2.8 minutes in nutrient-rich and of 4.6 min in starvation zones (Chen, Shiroguchi, et al., 2015a). Degradation was initiated for mRNA i as

$$Deg_{\mathrm{i,j,on|off}}(t) = \begin{cases} 1 & \text{if position of } RIB_{i,TL_{max}}\text{is} > 80 \\ 0 & \text{else} \end{cases} \tag{6.9}$$

with the movement of RNASE.

$$\frac{dx_{\mathrm{RNASE}}}{dt} = \begin{cases} v_{\mathrm{RNASE}} & \text{if initiated} \\ 0 & \text{else} \end{cases} \tag{6.10}$$

$$v_{RNASE,i} = \frac{L_{mRNA,i}}{t_{deg,med} - \frac{TL_{max} * \Delta x_{RIB}}{v_{RIB}}} \tag{6.11}$$

For transcripts longer than ≈ 1000 nts, degradation is initiated already when transcription is not finished, yet. Chen, Shiroguchi, et al. (2015a) found that 88 of 263 mRNAs showed lifetimes of the 5′-cap shorter than synthesis time of the transcript. Accordingly, co-transcriptional degradation was considered for long transcripts. Protein degradation is described using a constant rate degradation rate k_{deg} for the bulk proteins (Maurizi, 1992b). First order degradation kinetics were assumed depending on the nutrient condition, as

$$k_{\text{deg}} = \begin{cases} 0.01 & h^{-1} \text{ in nutrient rich zones} \\ 0.08 & h^{-1} \text{ in starvation zones} \end{cases} \tag{6.12}$$

Consequently, bulk protein *protein* of a subsequent time step *t+1* equals

$$protein\,(t+1) = protein\,(t) * (1 - k_{\text{deg}}) \tag{6.13}$$

6.3.10 Geometry and Reactor Setup

In order to consider a relevant industrial fed batch fermentation scenario a 54 m^3 stirred tank bioreactor was chosen. The main geometry was derived from Haringa, Tang, Deshmukh, et al. (2016) with precise dimensions and information about the inner geometry from Kuschel, Siebler, et al. (2017). The reactor setup included four baffles and a stirrer with two Rushton stirrers equipped with eight blades at the bottom and six blades at the top. With a stirring rate of 100 rpm the tip speed of 6.75 m s^{-1} was reached. The impeller Reynolds number was 2.77 × 10^6 and the required power was 225.69 kW, equaling a power number of 13.64. The feeding rate was set to 3.68 kg m^{-3} s^{-1} for an average growth rate of 0.2 h^{-1}. Aeration, gas transfer, and oxygen uptake were neglected in the study. The simplifying focus on the mono-phase conditions mirrors the basic strategy to showcase the propagation of transcript dynamics and the occurrence of additional large-scale ATP demands. Noteworthy, experimental data of Löffler, Simen, Jäger, et al. (2016b) were measured particularly excluding any impacts of oxygen limitation. Furthermore, said power inputs and cultivation conditions were chosen such that oxygen limitation is unlikely in the large scale scenario. Cell concentration of 31.8 $kg_{CDW} \cdot m^{-3}$ was assumed and a simple Monod-like kinetic was used to simulate the substrate uptake q_s:

$$q_S = q_{S,\max} \frac{c_S}{K_S + c_S} \tag{6.14}$$

where $q_{s,max}$ is the maximum uptake rate, c_s is the glucose concentration, and the approximated substrate specific uptake constant Ks with 4 mg L^{-1}. The maximum uptake rate was calculated with the biomass substrate yield Y_{XS} = 0.25 $g_s g^{-1}_{CDW}$ and the maximum growth rate μ = 0.2 h^{-1} (Villadsen et al., 2011). Based on the experimental observations in Löffler, Simen, Jäger, et al. (2016b) we concluded that stringent response is the predominant regulatory scheme initiated by repeated starvation. As a key characteristic stringent response reduces ATP consuming procedures trying to keep carbon supply on the maximum level achievable under stress conditions. Accordingly, we consider glucose uptake as a Monod-type function not being affected by the stringent response observed.

6.3.11 Simulation Setup

For the numerical simulation, the commercial calculation tool ANSYS Fluent version 19.1 was used. With 872,232 hexahedral numerical cells resulting in an aspect ratio of 12.6 and a minimal orthogonal quality of 0.34 high mesh quality was achieved. Schmidt number (Sc) tuning with Sc = 0.2 lead to the same circulation time as achieved by Haringa, Tang, Deshmukh, et al. (2016). The flow field was approximated by solving the Reynolds-averaged Navier-Stokes (RANS) equations in combination with the realizable k-ϵ model for turbulence. All surfaces were set as no-slip boundaries except for the no-shear top area which equaled the reactor filling height. Baffles and impellers were modeled as 0-thickness walls. Both impeller units were set to sliding mesh motion to generate a more realistic flow field. For glucose feed, a separate volume at the top of the reactor was defined, and a constant mass flow was set. The flow field and uptake kinetics were calculated every 10 ms until the glucose concentration was constant and a pseudo stationary gradient was reached showing constant metabolic activity. The conditions were 'frozen' for 180 s to track bacterial movements. These lifelines were simulated as massless Lagrangian particles with a discrete random walk (DRW) model passing through the flow field. Every 30 ms, position and glucose concentration of each bacterium were recorded. In total, 120,000 bacterial cells were tracked over around 180 s (residence time distribution: Appendix Figure D.6). According to the ergodic theorem, the same average values are obtained by tracking 1,080,000 bacteria for 20 s (for more information see Appendix: Supporting information F). However, due to limitation of simulation time and capacity

the simulation results were extended by repeating the single lifelines (Figure 6.2, B) every 180 s, while preserving the lifeline cluster groups. RNAP activities, mRNA levels, ribosomal activities and protein formation were calculated as described in chapter 3.4 considering each lifeline cluster (RNAP and mRNA profile of one lifeline group: Figure 6.2, C; mRNA content of individual cells after 180s: Figure 6.7, C). Finally, additional ATP demands were estimated (Figure 6.2, D).

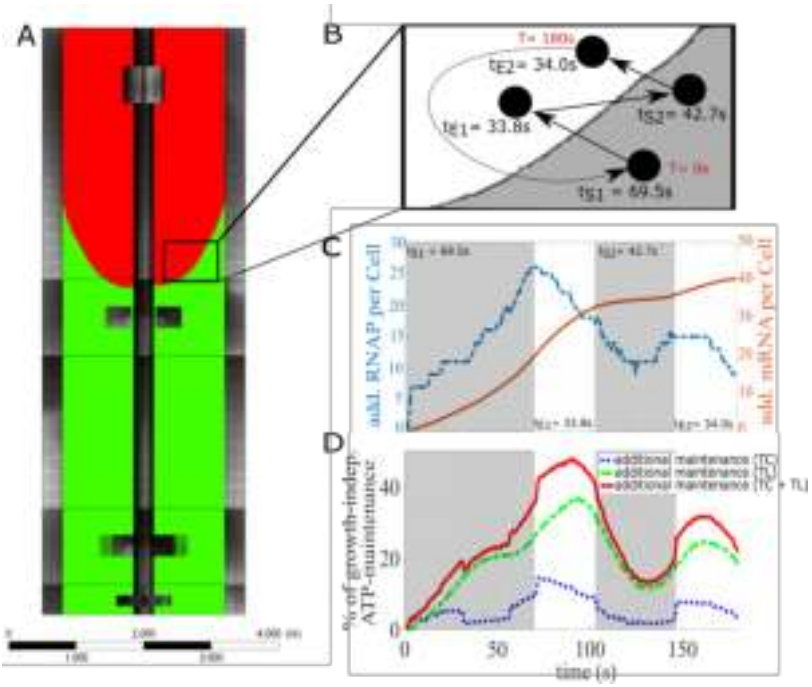


FIGURE 6.2: Impact of frequent exposure to feast and famine conditions in a large scale bioreactor (A). White are-as reflect nutrient excess while gray areas indicate starvation. The size of the areas reflect the corresponding residence time indicated with t_S for starvation and t_E for excess residence time (C: Bar plot for one cluster of particle trajectories). The starvation induced regulatory responses are propagated into the glucose excess zone causing a maximum growth independent ATP-maintenance in the glucose excess regime (D) based on additional active RNAP for transcription (TC) (C) and Ribosomes for translation (TL).

6.4 Results

6.4.1 Simulation results biological model

To identify data driven parameters for the model of *E. coli* K12 W3110, clusters of mean mRNA levels were identified (Appendix section C). Figures 6.3, 6.4, and 6.5 show the simulations (blue lines) and mean experimental values (red dots) for mRNA levels, active RNAPs, and the number of translated proteins per cell during PFR passage. The synthesis rate accelerates over time as more RNAP molecules are involved in the transcriptional response to the starvation stimulus. After a transcript is completely synthesized and the RNAP is released the number of active RNAP shortly drops before it rises again. Synthesized proteins appear with a delay that corresponds to the required translation time. Activities for transcription and translation result in additional ATP demands which are indicated as add-ons to non-growth dependent maintenance (NAM), shown in Appendix figure D.1. For transcription, costs are derived from nucleotide balancing including the release of nucleotides by mRNA degradation and the need for mRNA synthesis. Cost for translation mirror the amino acid needs and integration according to ribosomal activity. As indicated, translation costs are more than 2.5 fold higher than those of transcription. At maximum, cells have to bear 36.8 % additional NAM, 10.4 % coding for transcription, 26.4 % for translation. This happens during the early phase of frequent starvation exposure, i.e. after 3 starvation passages (25 min process time). After two hours process time, the ATP demand still increase. More than 45 % NAM increase is observed illustrating the remaining high number of active RNAP. Later, after 28 hours, NAM add-ons reduce more than five-fold compared to maximum needs. Then, transcription accounts for about 1 % NAM rise only. Total NAM increase only mirrors 9.5 %.

6.4.2 Linking cluster kinetics

Cellular adaptations to frequent environmental stimuli are mirrored in the cluster dynamics of differentially expressed genes (DEGs) that were measured after 25 min, 2h, and 28 h. Only 81 of 521 DEGs are conserved over the entire process time. This reflects the replacement of the initial sigma factor 70 dominated response by σ_{38} mediated regulatory programs (Löffler, Simen, Jäger, et al., 2016b). To simulate the transition so-called 'damping' and 'amplification' factors were identified using mean gene expressions as reference based on the simulated $\log_2$FC mRNA dynamics in Figure 6.3, 6.4, 6.5. Clusters were subdivided in 'persisting', 'subsiding' and 'non-active' fractions. The first collected genes

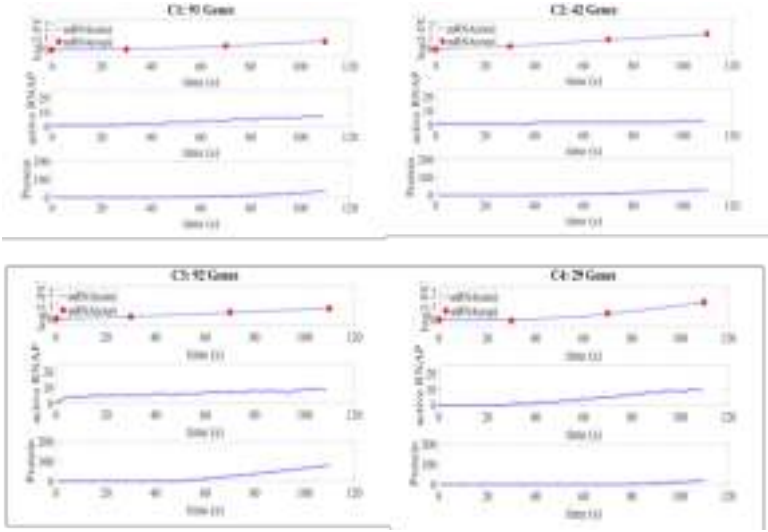


FIGURE 6.3: Simulated, cell-specific number of additional mRNA-levels (red dots: experiments; blue line: simulation), active RNAP and translated proteins of clusters 1 to 4 (C1-C4) along starvation passage (t=0-110s, 6.1) at 25 min. The logarithmic fold change (log_2FC) relates to STR values at the same process time.

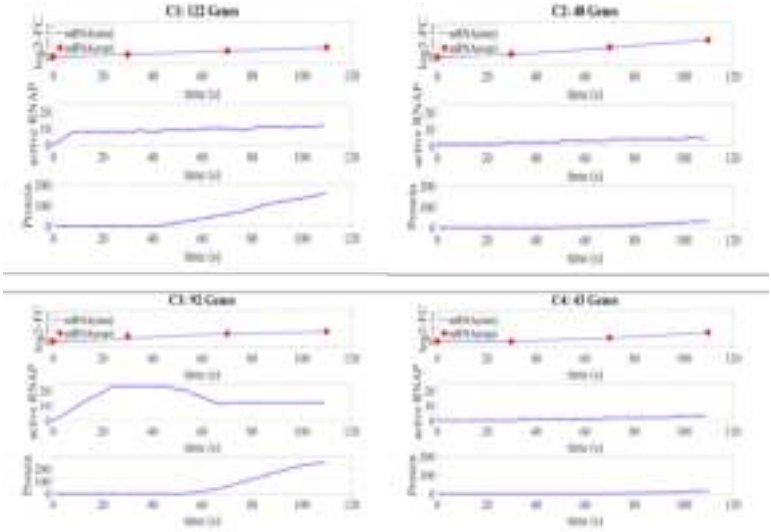


FIGURE 6.4: Simulated, cell-specific number of additional mRNA-levels (red dots: experiments; blue line: simulation), active RNAP and translated proteins of clusters 1 to 4 (C1-C4) along starvation passage (t=0-110s, figure 6.1) at 2 h. The logarithmic fold change (log_2FC) relates to STR values at the same process time.

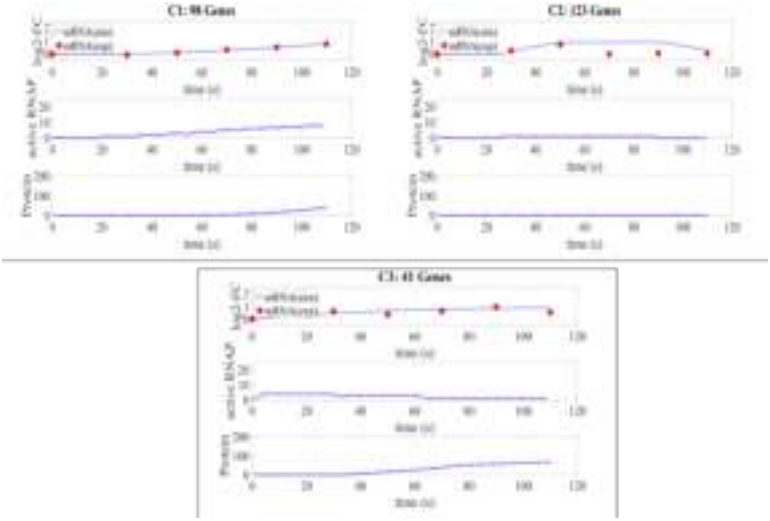


FIGURE 6.5: Simulated, cell-specific number of additional mRNA-levels (red dots: experiments; blue line: simulation), active RNAP and translated proteins of clusters 1 to 3 (C1-C3) along starvation passage (t=0-110s, figure 6.1) at 28 h. The logarithmic fold change (log_2FC) relates to STR values at the same process time.

with continuing high expression levels whereas the second comprised genes with declining expression levels. The last summed those genes that were either not yet or no more expressed between subsequent time points (Figure 6.6). The damping factor is the ratio of the mean log_2FC of subsiding and persisting genes for each cluster between two time points. The amplification factor is the ratio of the mean log_2FC of genes at 25 min which are active at time point 2h divided by the mean log_2FC ratio of genes activated after 2h (see exemplary calculation figure 6.6). The factors are used to calculate the amount of active RNAP:

$$x_{RNAP}(s,t) = x_{RNAP}(t) * (1 - \frac{Damping\ factor}{(S_{crit} - S_{min})} * (S(t) - S_{min})) \tag{6.15}$$

$$x_{RNAP}(s,t) = x_{RNAP}(t) * \frac{Amplification\ factor}{(S_{crit} - S_{min})} * (S(t) - S_{min}) \tag{6.16}$$

The entire transition process is guided by the number of starvation passages *S(t)* per time. S_{min} encodes the minimum and S_{crit} the critical number of passages. Whereas the first is a regression parameter, the later reflects experimental observations of Löffler, Simen, Jäger, et al. (2016b) as follows: 25 min equal 3 PFR (starvation) passages, 2 h equal 14, and 28 h equal 176. Noteworthy, the 28 h benchmark is chosen as a new steady state was observed already then (Löffler, Simen, Jäger, et al., 2016b). The modeling approach allows to transfer the STR-PFR observations to other conditions using the frequency of feast/famine exposure *S(t)* as key criterion.

6.4.3 Numerical Simulation

Glucose gradient

Applying the criterion of converged turbulent dissipation rate/power input the pseu- dostationary glucose gradient of figure 6.7 was obtained (figure 6.7). Accordingly, no further changes of glucose concentrations simulated at five locations occurred. The average concentration in the bioreactor was 23.74 mg L^{-1}. For comparison, the average glucose level observed by the Lagrangian particles (cells) was 22.79 mg L^{-1}. Consequently, only 4 % deviation was found which is qualified as a small difference indicating good homogeneous distribution and reflecting impacts of the turbulence model and of particle lifeline filtering. The volumetric distribution between starvation and excess zone is 73 % to 27 %, respectively. Again, similar percentages were calculated by integrating mean residence times of all lifelines. The mean residence time of the cells in the starvation regime is 9.46 s

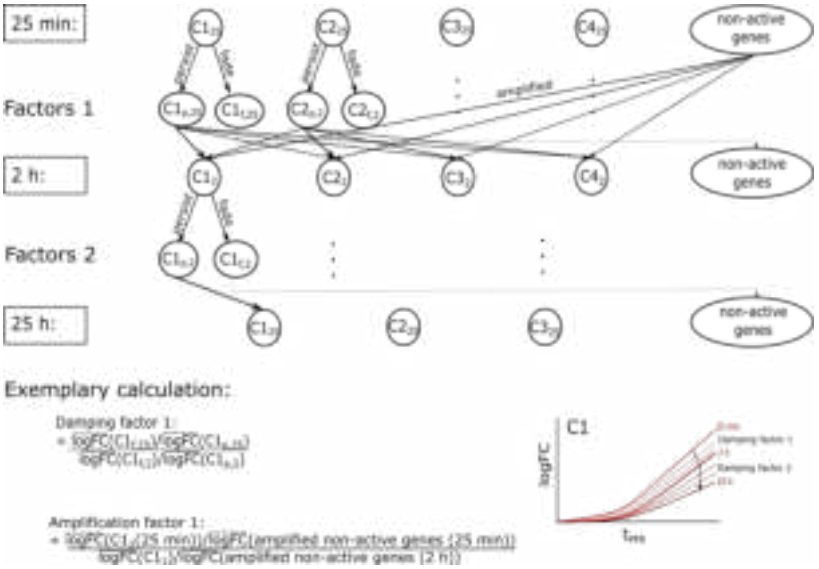


FIGURE 6.6: Scheme illustrating how 'damping' and 'amplification' factors are derived from experimental cluster da-ta (C1-C4) measured at 25min, 2h, and 28h after PFR passage. In general, active genes of individual clusters may continue amplification (persisting, p), reduce expression (fading: f), or even be activated from the group of non-active genes. To bridge the gene expression dynamics from 25 min to 2 h and from 2 h to 28 h, so called damping and amplification factors are calculated as illustrated in the example. They use mean logFC values of the relevant time points. The damping factor correlates the ratios of fading-to-persisting genes of the two subsequent measurements. By analogy, the amplification factor calculates ratios of mean numbers of gene expression versus 'first time' amplified genes for each time point and correlates the same for two subsequent events.

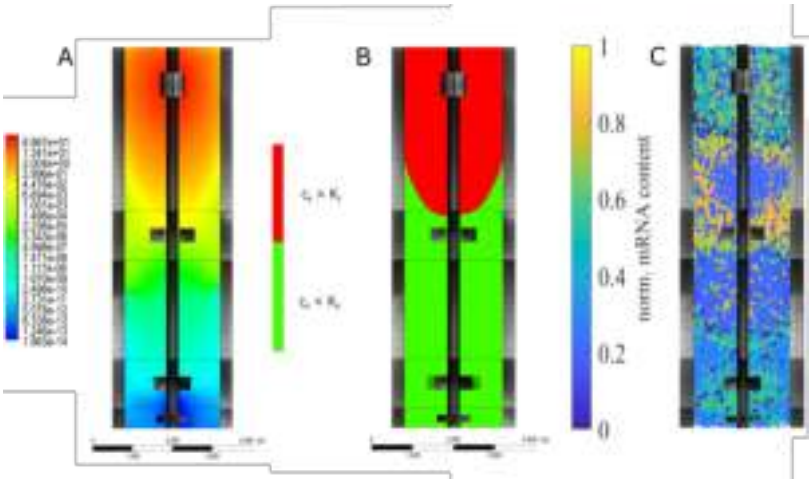


FIGURE 6.7: Simulation results of the 54m³ stirred tank reactor. A, Left: Log-contours of cs/Ks gradients. B, Middle: Assignment to regimes (limitation, green: cs < Ks; saturation, red: cs>Ks); C, Right: normalized mRNA content analyzed after cells fluctuated 180 s through the bioreactor

(Appendix Figure D.6) which is in the same range as published by Haringa, Tang, Deshmukh, et al. (2016).

Ideally, large-scale simulations should have been compared with real in situ measurements to challenge the predictions. However, such data are missing, which represents a common problem often faced by academic groups. Nevertheless, applying CFD simulations still offers best chances for getting highly accurate large-scale predictions as complex hydrodynamics even including overlapping flow fields between stirrers are well predictable. Notably, the latter may hamper the application of simplifying compartment-based estimations which basically assume separated flow fields between stirrers (see Appendix D.11).

For simplicity, Eulerian simulations only considered the liquid phase thereby assuming sufficient oxygen supply in the bioreactor without calculating dissolved oxygen concentrations explicitly. Furthermore, additional turbulence due to bubble interaction was neglected, too. Substrate consumption followed Monod-type kinetics taking place in each numerical cell. This implied that bacterial cells were homogeneously distributed at each time step.

Statistical Evaluation

Lifeline statistics reflect the imprint of changing micro-environmental conditions on cells fluctuating through the bioreactor. To be precise, cellular residence times in different concentration zones and shifts between proximal regimes were studied. At start, cells were 'inserted' into the bioreactor along a straight line reaching from top to bottom. After a few simulation steps, cells were distributed homogeneously before individual cell tracking started for 180 s. Lifeline records were cured by percolating only those with residence times longer than 0.13 s. The latter represent unrealistically turbulent fluctuations. The following threshold was defined for regime analysis: If cells experience lower or higher glucose levels than K_S for at least one second, the period is labelled as starvation or saturation time, respectively. Noteworthy, the minimum residence time of one second equals the average metabolite turnover time in *E. coli* (Taymaz-Nikerel, Van Gulik, et al., 2011; Shamir et al., 2016). The implementation of the harsh regime boundary K_S finally lead to rapid and somewhat artificial regime shifts. They were excluded from analysis by ignoring the upper and lower 1 % of regime changes. Alternately, the consideration of alarmones such as (p)ppGpp serving as intracellular triggers to initiate transcriptional regulation may yield at continuous models. Unfortunately, understanding of alarmone formulation, degradation, and alarmone induced regulation is still too fragmented to build dynamic transcriptional models. In total, measures for residence time percolation and shift filtering only reduced the data set by 3 % (Appendix figure D.6).

At maximum, 41 regime shifts were observed during the 180 seconds observation period. Most frequently, 20 regime changes occurred and cells rested in single zone no longer than 30 seconds. As a key characteristic, cells once exposed to glucose starvation reset their regulation signal. But RNAP and ribosomes remain active propagating the starvation response into the glucose excess regime (Figure 6.2, C, D). According to their starvation pattern the cell lifelines were assigned to 70 different clusters. Thereby each cluster represents a specific fluctuation pattern reflecting the amount and duration of changes between starvation and excess zone (Figure 6.2, B).

6.4.4 Coupling the biological model with lifelines

To minimize the computational efforts particle lifelines were exported from ANSYS Fluent. Starvation patterns of lifelines (Figure 6.2, B) served as input for the biological model. A workflow scheme is provided in figure D.8, appendix. It was assumed that each entry of the starvation zone activated the expression of distinct gene clusters as experimentally

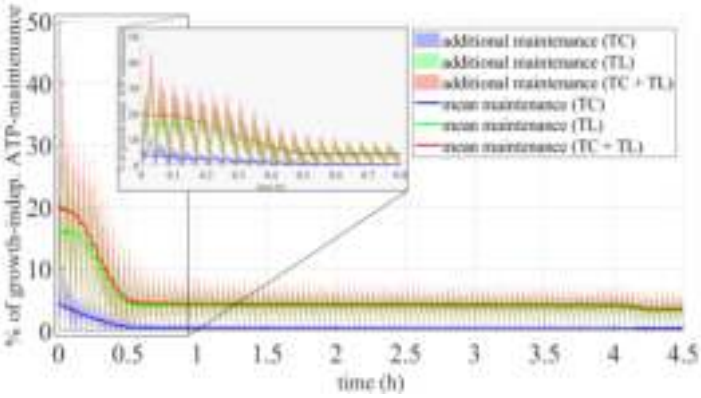


FIGURE 6.8: Additional ATP demands of a population of 120,000 'newborn', not preconditioned cells in a 54 m3 re-actor over 4.5 hours process time. Courses for mean transcription (blue: TC), translation (green: TL) and the sum of both are depicted (red: TC+TL). The shaded areas display the standard deviation.

observed. Accordingly, individual expression patterns were estimated for each cell mirroring their particular tracking history. As indicated in figure 6.7C, the basic expression level of the cellular population is increased by 37.7 % compared to the reference. This reflects the additional cellular needs to adapt to the heterogeneous mixing conditions in the reactor. Noteworthy, high mRNA levels, induced by preceding starvation, are propagated into glucose rich zones (Figure 6.7, C). Expression patterns reflecting starving conditions occur in saturating glucose zones and *vice versa*. The phenomenon mirrors the delayed transcriptional response that is slower than convective movements of cells in the bioreactor. Consequently, a high transcriptional heterogeneity occurred in the tank. A maybe surprising pattern is disclosed: In the lower part of the reactor, cells envisage low glucose concentrations but show reduced mRNA levels (norm. mean mRNA level: 0.33). On contrary, cells facing high glucose levels in the upper part reveal high mRNA levels (norm. mean mRNA level: 0.42). About 25 % of the cells permanently stay in the starvation zone. This fraction even adapts to the limiting conditions which reduces the transcriptional response gradually. Cells located close to zones of glucose excess highly fluctuate between starving and saturating conditions. Consequently, strong gene expression responses are observed. The average ATP-demand of a newborn, not pre-conditioned population of 120,000 cells exposed to the 54 m^3 bioreactor is shown in Figure 6.8. Basically, plot 8 illustrates the cyclic passing of 180s lasting lifelines. To filter related peaks only average

values are indicated using a moving median filter over 700 data points. Furthermore, a moving standard deviation with a window size of 400 data points is added as shadow. Synchronization-like patterns reflect clustering of particles in groups. 70 bins were used with passable computational effort. The maximum of 45 % additional maintenance is predicted shortly (0.03 h) after cells were exposed to the bioreactor condition. After about 0.5 h most of the population has adapted to the heterogeneous environment reducing the additional ATP demand to 6.5%. After 4.22 hours the last cell fraction is in adaption state.

6.5 Discussion

To disclose spatial heterogeneities of regulation patterns and additional ATP demands of *E. coli* K12 W3110 exposed to a 54m³ stirred tank reactor CFD based lifeline analysis was coupled with agent-based modeling for transcription and translation. As a prerequisite, a proper model describing stress induced dynamics of transcription and translation is needed. Applying a clustering approach, it was possible to properly describe the experimentally observed transcription dynamics (Löffler, Simen, Jäger, et al., 2016b) of 821 genes using 16 parameters. Under steady-state conditions newly synthesized and recycled nts equilibrate in cells before they enter PFR. However, ATP demands for transcription rise inside PFR as mRNA synthesis exceeds the recycling rate. The introduction of 'amplification' and 'damping' factors managed to model the transition from fast response to long-term adaptation, as visible in the experimental data. Accordingly, modelling succeeded to mirror the cellular efforts shifting control from σ_{70} to σ_{38}, more and more (Löffler, Simen, Jäger, et al., 2016b). Given that *E. coli* cells with doubling times of 3.3 hours contain about 120 active RNAPs (Bremer et al., 2008), approximately one fifth of the available RNAPs at 25 min is used in the transcriptional response (Figure 6.3). The amount of involved RNAPs slowed down after 28 h for two reasons. First, the absolute number of initiated transcripts is reduced which mirrors cellular adaptation. Second, transcription even stopped after 50 s for a large group of genes (Figure 6.5). Accordingly, reduced synthesis costs occur. This is amplified by the prolonged life time of transcripts during starvation which further reduces the amount of synthesis to obtain a certain level of mRNA abundance as described in chapter 3.4.4. mRNA lifetime is proportional to their distance from the 5′ end of the transcript according to (Chen, Shiroguchi, et al., 2015a), what is in line with the observation that the 5′ end contains important determinants that regulate RNA lifetime (Arnold et al., 1998). Appendix figure D.1 shows that protein synthesis accounts for the major part of ATP consumption. Peak values reach about 45 %

NAM at 2 h process time. Later on, the demand steadily reduces as less mRNA transcripts are synthesized and less proteins are translated. At 25 min, the number of active ribosomes involved in the stress response rises steadily during starvation reaching 10 % of all available ribosomes (Reference: 10^3 ribosomes/cell, (Bremer et al., 2008). The fraction decreases to 5 % after 28 h reflecting the adaptation process. Linking the transcription and translation model to large-scale lifelines reveals the impact of delayed cellular response to fast external changes. In essence, cellular responses of transcription and translation are slower than convective zone shifts. Consequently, they are transported from one location to another, basically decoupled from external changes. Spatial analysis of all cells after 180 s (Figure 6.7, C) reveals highest transcript levels close to or even inside the glucose excess regime (50-80 % mRNA upregulation) whereas the lowest are found at the bottom. Once initiated, the starvation response propagated into the glucose excess zone. There, additional needs for transcription, translation, and ATP may limit the targeted formation of industrial products in microbial cells. Noteworthy, it is exactly this feature that distinguishes the current model from previous lifeline studies (Haringa, Tang, Deshmukh, et al., 2016; Haringa, Deshmukh, et al., 2017; Kuschel, Siebler, et al., 2017). There, metabolic and cell cycle responses were considered as *instantaneous* cellular response. Fast external changes are immediately translated in cellular replies. Later, Haringa, Tang, Wang, et al. (2018) implemented metabolically buffered responses by considering variable enzyme pools (Tang, Deshmukh, et al., 2017). In this context, the current study proceeds by additionally integrating downstream transcriptional responses incorporating another level of cellular control. Our approach introduces the *non-instantaneous*, delayed response by considering intracellular programs of longer time scales than external changes. Accordingly, responses may propagate in different zones of the reactor causing non-expected transcriptional regulation programs, there.

The approach was exploited further by estimating the entire add-on ATP demand for 120,000 newborn cells monitoring 4.5 hours process time (Figure 6.8). As shown, the adaptation of the population is finished after 4.2 h disclosing a remaining ATP add-on of about 6 % NAM compared to the 45 % max NAM at the beginning. In terms of microbial productivity, these ATP needs simply reduce the amount of available ATP for product formation, i.e. they limit biomass specific productivities. The phenomenon has often be described in large scale fermentations (Lara, Galindo, et al., 2006). Noteworthy, it is likely to be pronounced in hyper-producing cells with ATP intensive product formation. Often enough, such production processes run in fed-batch mode installing reduced, limiting metabolic activity to stay within the technical limits of aeration and cooling. Consequently, those

additional ATP needs hit cells with limited ATP forming capacities.

To evaluate the impact of particle simulation time an additional simulation was conducted using a high performance computation cluster studying 60,000 particles for about 460s. Similar results were obtained for the key readouts, e.g. time courses of ATP maintenance demands and residence time distributions remained. The simulated adaptation time reduced from 4.2 h to 3.7 h mirroring the lowered amount of particles staying in the starvation zone for the entire process time (around 5 %) (see supporting information J). However, modelers need to consider that long simulation times are likely to violate the intrinsic constraint of one-way coupling neglecting particle-environment interactions for the sake of simplicity. In this dilemma, we decided for the analysis of 180s to capture key dynamics while still fulfilling the one-way coupling constraint.

As pointed out by Löffler, Simen, Jäger, et al. (2016b), the majority of transcription dynamics is caused by the frequent on/off switching of stringent response, mediated by rising intracellular (p)ppGpp levels. Hence, creating stringent response deficient strains (Michalowski et al., 2017) opens the door to prevent non-wanted NAM increase. Besides, other cellular stress programs may be targeted as well (Appendix: Table D.4).

6.6 Conclusion

The current modeling approach marries computational lifeline analysis with cellular regulation models thereby introducing the *non-instantaneous* cellular response to changing extracellular conditions. Consequently, the spot of stress induction and the location of cellular phenotype do not need to be the same. Accordingly, heterogeneities in large-scale bioreactors comprise the physical levels linking local conditions tightly with metabolic responses and the cellular regulation level encompassing delayed responses such as transcriptional or translational effects. To detect the latter and to describe them properly in data driven models experimental scale-up simulators are necessary that mirror transcriptional and translational cellular replies as performed by Kuschel and Takors (2020). The setting of such devices may differ from 'conventional' scale-up simulators that typically mimic circulation times. Because the entire transcriptional responses should be clearly detectable rather long stress exposure periods should be installed and read.

Chapter 7

Conclusion

An exponentially growing human world population with an increasing need for more and cheaper sustainable products presents a huge challenge for industrial biotechnology. Economically efficient processes and production capacities must be exploited and optimized by tackling scale-up challenges, to enable efficient mass production. The complexity of bioreactor scale-up is a consequence of the interplay of interdisciplinary challenges, such as robust strain design and optimal process performance, as displayed in figure 7.1.

Understanding the function of cellular behavior under varying conditions requires the development of computational approaches that incorporate gene regulatory models as well as environmental perturbation simulations that rely on experimental evidence. In this thesis a method is provided for achieving this goal, by combining the two approaches. This chapter places the individual research topics into a scientific context by answering the research questions listed in 1.

Answering the Research Questions

If a newly engineered strain optimized for large-scale bioprocesses is developed, the behavior of the altered microbe must be the result of consistently engineered changes, rather than spontaneously introduced mutations that are due to stress. One way to check for inconsistencies is presented in the answer to question 1.

> **A1**: Less costly than RNA-Seq, DNA-Seq is suitable for the identification of structural variations at the end of a series of experiments, to evaluate the newly engineered strain. Unicycler, Mauve and Snippy serve as the main assembly and analysis algorithms.

As presented in chapter 3, the sequence comparison of both *E. coli* strains (mutant and parental) revealed presence of sequence variations, but no problematic mutations were

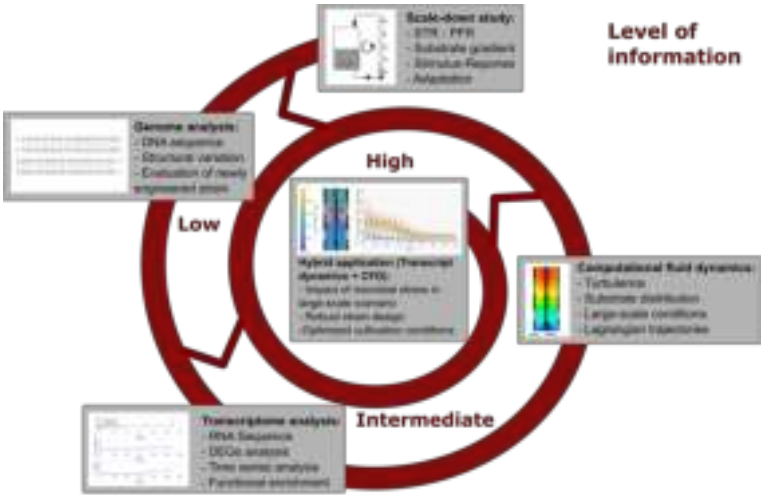


FIGURE 7.1: Summary of experimental and numerical simulation approaches leading to a hybrid approach which combines transcript dynamics with large-scale CFD simulation

identified, and there was no influence on the favorable results that were observed. Consequently, the genomic changes that lead to a significantly lower maintenance coefficient and increased robustness with regard to the influence of heterogeneities can be traced back to the engineered robust strain chassis, rather than to stress-induced mutations.

To determine the impact of heterogeneous process conditions and evaluate the performance of an engineered strain (*E. coli* SR) optimized for large-scale fermentation processes, in comparison with its parent strain (*E. coli* K-12 MG1655), both strains were cultivated in a scale-down device in which nitrogen starvation was repeatedly induced (see chapter 5). Transcriptomic analysis was used to derive the transcriptional landscape of both strains and led to the answer to research question 2.

> **A2**: The core algorithm used in the established workflow (DeSeq2) was suitable for identifying up to 837 statistically significant up- and down-regulated genes, in addition to regulation pathways for the identification of mechanisms for coping with fluctuating substrate conditions, as well as loss- and gain-of-function mutants associated with the mutant phenotype.

Differential gene expression analysis yields information on the complete set of mRNA transcripts that are produced under specific circumstances, due to their direct link to the cell's environment. Based on the workflow established, the regulatory response of microorganisms can be revealed, to optimize the production host by identifying strategies and bottlenecks that arise during microbial adaptation.

Analysis of mRNA transcripts in chapter 5 following repeated nitrogen starvation in a STR-PFR scale-down system, revealed clearly distinguishable transcriptomic responses for the two *E. coli* strains, which led to the answer to question 3.

> **A3**: Compared with *E. coli* MG1655, the mutant SR strain displayed better suitability to large-scale bioprocesses, due to a diminished reaction towards repeated nitrogen starvation.

Basically, *E. coli* SR only displays Ntr response to nitrogen starvation, and no ppGpp induced response was shown. However, *E. coli* SR seems to counteract the missing ppGpp response with partial cell death, negative regulation and induction of toxin/ antitoxin systems to restrict transcription and translation processes. Although it was possible to identify certain mechanisms in the SR response to repeated nitrogen starvation, the complex response as a whole, and the various interactions that take place, are not yet fully understood.

As already indicated by the answer to question 3, transcriptional regulation alone may be insufficient to describe stress-induced transcription changes. Other cellular processes such as signaling, metabolomics or proteomics must be monitored at the same time. Even though no proteomic data were gathered in this study, translation is inherently coupled to transcription, since energy preservation precludes the accumulation of unused transcripts. Therefore, translational cost predictions regarding protein formation are essential, in order to complement the regulatory model. The answer to question 4 provides an extension of the model predictions.

> **A4**: The molecular translational process can be modeled in a co-transcriptional manner substantiated by values from the literature. In the process, ribosomes are included as a limited quantity with a translational elongation rate equal to the transcriptional elongation rate. A maximum of 11 translations per mRNA strand inhibits the unlimited production of protein.

Experimental evidence can thereby validate and complement or correct certain assumptions in the future. To make the analysis outputs accessible to a large group of researchers, a comprehensive, yet simple expression is necessary, which sums up the transcriptional and translational changes observed. This challenge led to the answer to research question 5:

> **A5**: ATP requirements for the formation of amino acids and nucleotides were calculated. The overall costs of amino acid synthesis and polymerization were estimated as 4 ATPs, whereas the absolute numbers of synthesized and degraded nucleotides were derived from experimental data. For recycling, nucleotide costs of 2 ATP were assumed. As a reference, growth independent maintenance (NGAM) was used, at 0.0027 mol ($g_{DW}\,h^{-1}$).

Based on the work of Kaleta et al. (2013), every cellular process was broken down into either anabolic (ATP demanding) or catabolic (ATP recycling/ production) reactions, to calculate the overall cost of a molecule in terms of ATP. Since cells have an upper limit of ATP production capacity, as a higher level of stress usually results in slower growth and limited productivity, this measure can be used to translate stress-induced transcriptomic changes into add-on to NGAM demands, permitting the visualization of potential limitations in cellular production capacity.

After certain regulation patterns in *E. coli* were identified, the question arose as to which model would be suitable for displaying the observed dynamics. The examination of simulation methodologies in section 2.6 led to the answer of research question 6.

> **A6**: Continuous models, such as ODE models, are suitable for displaying the dynamic regulatory changes that were observed. They can quickly provide a detailed representation of the quantitative changes, and do not require huge computational capacity. When these models are combined with stochastic elements, it is possible to represent the behavior of individual molecules in a more realistic way, than the ODE approach alone.

Due to their reliance on a very few variables, Boolean models yield only a very abstract view of qualitative dynamic behavior; their main advantage is the speed at which calculations can be made. At the same time, single-molecule level models are not the only suitable models for this purpose. While they take into account the full complexity of gene

regulation, single-molecule level models are the hardest to study analytically, and the most expensive when it comes to computational analysis. Also, stochastic experimental data is currently very hard to find. For these reasons, an agent-based model combining ODE and stochastic elements was used in this study. In an approach that picks up on the work of Nieß et al. (2017), members (e.g. RNAP, ribosomes) are treated as autonomous molecules that are characterized by certain properties, in accordance with an ODE formalism. That means that, all molecules are represented by discrete values in accordance with observations made during experiments, and differentially expressed genes are randomly chosen for expression. This implementation offers the potential for uncomplicated adaptation to other conditions or organisms, as long as organism-specific parameters are available. Additionally, a rather simple mathematical formulation makes it possible to extend the model, including the metabolic response, to a multi-layer system.

To determine the biological response in an industrially relevant scenario, a realistic method must be used to investigate the cellular environment inside an industrial-scale bioreactor. A single-phase fluid flow simulation in CFD was used to simulate substrate distribution in a 54 m^{-3} reactor. However, there is a large number of different approaches to displaying this heterogenic environment, which provides the reason for answering question 7.

> **A7**: The Euler-Lagrange modeling approach based on the realizable k-ϵ turbulence model, in combination with SM, was suitable for simulating a glucose gradient in a stirred tank reactor.

After the simulation of a glucose gradient and particle lifelines via a Euler - Lagrange approach, it was necessary to combine the CFD results with the biological model. Due to computational limitations, it was not possible to establish a direct link between the agent-based model and the Euler-Lagrange framework. That means that the continuous resolution of the biological model and its impact on the reactor environment (which could yield a more realistic representation than otherwise possible) could not be included in this study. Consequently, assumptions have been made to simplify the modeling process, as summed up in the answer to question 8.

> **A8**: Two computationally intensive analysis methods were combined, by using Lagrangian trajectories to record particle location and glucose substrate concentration, and employing agent-based modeling to predict the biological response to it in a

Matlab environment. Starvation duration and frequency as identified via particle tracks were used as input parameters in the biological model.

The key question in this study arose after both models were combined, and the consequences of insufficient mixing for microbial energy levels were investigated. The answer to this question follows:

A9: In order to achieve starvation adaptation in a large-scale reactor, *E. coli* needs up to a maximum of 45 % in additional growth-decoupled maintenance shortly (0.03 h) after the cells have been exposed to the bioreactor condition. After about 0.5 h, most of the population has adapted to the heterogeneous environment, reducing the additional ATP demand to 6.5 %. After 4.2 h, the last cell fraction is in an adaptation state. These results indicate delayed and reduced production of ATP-intensive products, ultimately resulting in a reduced space-time-yield, and loss of economical attractiveness.

The research questions answered so far conclude the main part of this thesis. The remaining research questions are concerned with the extension from single to multiphase conditions, including the safflower oil experiments presented in Appendix E. Since the exact distribution of safflower oil in a turbulent system and the interaction with bubbles is unknown, preliminary studies were performed with a main focus on bubble surface tension, size distribution and oxygen mass transfer coefficients in the absence and presence of safflower oil. The first question, therefore, refers to experiments performed to investigate the mentioned parameters, and it is followed by the implementation strategy for a simple oil-in-water system in a CFD environment, in order to examine oil droplet distribution in a large-scale scenario.

A10: Safflower oil reduces surface tension of bubbles in water from 71 mN m^{-1} to around 61 mN m^{-1} at thermodynamic equilibrium, but it showed no effect in a complex glucose medium, as surface tension was already lowered to around 35 mN m^{-1}, mainly due to the presence of salts. Measurements of bubble size distribution showed reduced bubble diameters when oil is present. However, k_La measurements indicate the opposite of the anticipated effect.

Reduced bubble size lead to an increase of the interfacial area, potentially resulting in

increased species mass transfer. However, k_La values in the presence of safflower oil, regardless of the setup or medium used, were diminished by up to 50 %, compared with oil-free systems. This shows that oil has a much more complex effect on mass transfer than expected, which cannot solely be explained by an increase of the interfacial area. A thorough analysis is needed to quantify these effects.

By elucidating the putative contributions of modeling, the discussion presented above has highlighted the need for knowledge-based process scale-ups. In this thesis, a framework has been presented that includes a transcriptional analysis workflow to gain insight into the biological response of starvation-induced adaptation processes, which in turn has been translated into an agent-based regulatory model. When the agent-based model is combined with a glucose gradient simulation in a 54 m^3 bioreactor, they serve as a method for a facilitated *a priori* large-scale process performance test of robust strain design as well as optimized process conditions. In detail, the presented methodology enables the combination of scale-down experiments with numerical predictions of heterogeneous large-scale environments and their effect on microbial hosts in terms of additional growth-decoupled maintenance demand. Complex transcriptomic analysis results can be translated into an easy-to-interpret formulation with the use of ATP as a universal energy currency. Additionally, through the use of this method it is possible to predict the extent to which the overall process time in an industrial scale scenario will be increased due to microbial adaptation. In addition to transcriptomic analysis, genome analysis has proved to be a simple, cost-efficient and valuable tool for validating engineered strains in heterogeneous scenarios, and for supporting the design of other strains. Since the presented methodology is versatile, it can be extended in accordance with the top-down approach, by including results from further studies that display a range of granularity. Additionally, it can be adapted to study other stress factors and organisms.

Chapter 8

Outlook

This thesis provides numerical methods for the investigation of substrate distribution inside a large-scale bioreactor, as well as the biological response to it. However, the methods presented are subject to limitations. As microorganisms face depletion of multiple nutrients (e.g. nitrogen, phosphorus or oxygen) in a heterogenic bioreactor, a combination of nutrient starvation conditions as presented by Kuschel and Takors (2020) should be tested in STR-PFR experiments, and modeling approaches extended accordingly. Additionally, by-product formation might be an important factor to consider in accounting for media composition changes as presented in a recent study published by Sarkizi Shams Hajian et al. (2020). Since multi-component approaches are computationally intensive, CFD simulation is currently limited by the computing power available for the respective investigation, and often only snapshots over the range of a few minutes of the bioreactor environment including the microbial producers can be represented. As the technology develops further, simulations of more complex interactions over a longer period of time may be feasible.

Despite the current limitations of numerical simulations, further investigation is warranted into the contribution that they can make to *in vivo* experiments that incorporate large-scale conditions and single-cell resolution. But due to a lack of large-scale experimental data, many regulation theories are still based to some extent on empirical observations. The prediction accuracy of biological models can be improved by the further development of single-cell resolution techniques, which are a promising tool for increasing the understanding of differential gene expression in microorganisms (Brognaux et al. (2013) and Taheri-Araghi et al. (2016)). Especially when combined with comprehensive omic data analysis, including transcriptomic, proteomic and metabolomic data sets, biological models will help to identify complex regulatory networks. Further research should focus on combined approaches within a short sampling time to fully understand microbial behavior. Although work on this combined approach has just begun, it has significant

potential for further developments with regard to the design of reactors and strains.

However, one inherent problem of transcriptomic data analysis is the enormous variety of methods and algorithms, which makes possible subjective interpretations. This diversity pushes the scientific research forward, but at the same time makes it difficult to find an exact explanation for certain phenomena. Variations in the way experiments are carried out, and in sequencing methodology add to the problem.

In conclusion, models resemble reality only to a given degree. Many process parameters, such as temperature or pH, are not fully considered, and the complexity of an organism as a whole has yet to be represented. Each assumption that is made takes the model further away from reality. As Almquist et al. (2014) have stated, the real value of a model lies in using it to predict, evaluate, and explore different scenarios or assumptions involving the modeled system and its surrounding environment. An established model should thus be seen first and foremost as a tool that can be used to answer questions about the microbial host or process conditions, and it should be used as a complement or alternative to performing actual experiments.

Appendix A

Cluster evaluation results

TABLE A.1: Cluster indices at different time points for k-means clustering

Timepoint	Silhouette S	Figure of merit	Intra cluster variance
25 min	0.682	0.097	0.030
2 h	0.724	0.781	0.015
28 h	0.562	0.068	0.077

Appendix B

Engineering of robust *Escherichia coli* chassis strains for large-scale production processes: Supporting Information

Supporting information S1:List of all primers used

TABLE B.1: Primer oligos used in this study

No.	Name	Function	Sequence 5′ -> 3′
MZ_013	delta_fliR_downstream_fwd_(TS2f)	genomic deletion from fliE to fliR	TTCCGTAACGTTTATCATGTTAT
MZ_014	delta_fliR_downstream_rev_(TS2r)	genomic deletion from fliE to fliR	ATTTGAATGGTTCCCTGACCT
MZ_015	delta_fliE_downstream_fwd_(TS1f)	genomic deletion from fliE to fliR	AATTAGCCCAGTCGCTGAAA
MZ_016	delta_fliE_downstream_fwd_(TS1r)	genomic deletion from fliE to fliR	ATAACATGATAAACGTTACGGAAG
			TTTTGTAACCTGTTGTTAATTACA
MZ_033	Δ_flK_fwd_TS1f	genomic deletion of flk	CGCCAGATGTAGCGTTTTGT
MZ_034	delta_flK_fwd_TS1r	genomic deletion of flk	GATACTGTCATTTCGTCGGGGTACGT
			ATCCTTATACCTGAAATCTTC
MZ_035	delta_flK_fwd_TS2f	genomic deletion of flk	CCCCGACGAAATGACAGTATC
MZ_036	delta_flK_fwd_TS2r	genomic deletion of flk	GAATGGCGATTTACGGTGCC
MZ_055	flk_Deletionsanalyse_fwd	seq. of Δflk strains	GGTCAGCAACGCAGTATTATCG
MZ_056	flk_Deletionsanalyse_rev	seq. of Δflk strains	CTCTGGCAAAGTGATGTCATGG
MZ_083	fliA_Deletionsanalyse fwd	seq. of ΔfliA strains	GCCATCACACCCATCAATGC
MZ_084	fliA_Deletionsanalyse rev	seq. of ΔfliA strains	GTCAAACTGGGCGGAGATGA
MZ_085	deltafliA_TS1f	genomic deletion of fliA	CTGGAGGATTTCTGCACAAG
MZ_086	deltafliA_TS1r	genomic deletion of fliA	CGTCAGTAAATGCCGCAC
MZ_087	deltafliA_TS2	genomic deletion of fliA	GTGCGGCATTTACTGACGG
			ATAAACAGCCCTGCGTTATATG
MZ_088	deltafliA_TS2r	genomic deletion of fliA	CTGACTGCTGTTGCAAATGG
MZ_089	Tet-SacB deltafliA 1	genomic deletion of fliA	ATCATTAAGAACTCCTGGTAGTCAAAGT
			TAAAGTGCGGCATTTACTGACGTCCTAATT
			TTTGTTGACACTCTATC

No.	Name	Function	Sequence 5′ -> 3′
MZ_090	Tet-SacB deltafliA 2	genomic deletion of fliA	CAGAAACGGATAATCATGCCGATAACTCATATAACGCA GGGCTGTTTATCATCAAAGGGAAAACTGTCCATATGC
MZ_091	Tet-SacB fwd	amplifies tetA-sacB or fragments for SOEing PCR	TCCTAATTTTTGTTGACACTCTATC
MZ_092	Tet-SacB rev	amplifies tetA-sacB or fragments for SOEing PCR	ATCAAAGGGAAAACTGTCCATATGC
MZ_093	Tet-SacB deltaflk korrigiert fwd	genomic deletion of flk	AATCTCGCGGCCAGGCATACTTGCGAAGATTTC AGGTATAAGGATACGTATCCTAATTTTTGTTGACACTCTATC
MZ_094	Tet-SacB deltaflk korrigiert rev	genomic deletion of flk	ATCTCTGCGGTGCTGGGTATTATTGTCACGATACT GTCATTTCGTCGGGGATCAAAGGGAAAACTGTCCATATGC
MZ_095	Tet-SacB Soeing rev (Tet)	amplifies tetA-sacB or fragments for SOEing PCR	GTTTGAAGATGGCAAGTTAGTTAC
MZ_096	Tet-SacB Soeing fwd (SacB)	amplifies tetA-sacB or fragments for SOEing PCR	GTAACTAACTTGCCATCTTCAAAC
MZ_098	Tet-SacB deltafliC 2	genomic deletion of fliC	GGTGGAAACCCAATACGTAATCAACGACTTGCAATAT AGGATAACGAATCATCAAAGGGAAAACTGTCCATATGC
MZ_099	deltafliC_dflia_TS1f	genomic deletion of fliC (in a ΔfliA context)	AACGACATATTCACGCACC
MZ_100	deltafliC_dflia_TS1r	genomic deletion of fliC (in a ΔfliA context)	TCGTTGTAACCTGATTAACTGAG
MZ_101	deltafliC_TS1f	genomic deletion of fliC	GGCTGTTATTGGTGTCGAG
MZ_102	deltafliC_TS2f	genomic deletion of fliC	CTCAGTTAATCAGGTTACAACGA GATTCGTTATCCTATATTGCAAGTC
MZ_103	deltafliC_TS2r	genomic deletion of fliC	CCACGTTAATGATGCTTGC

No.	Name	Function	Sequence 5′ -> 3′
MZ_104	Deletionsanalyse fliC fwd	seq. of ΔfliC strains	GCCACTCATCGTAGGAGAAG
MZ_105	Deletionsanalyse fliC rev	seq. of ΔfliC strains	GATGTGACTGACAGACGATATTC
MZ_106	Deletionsanalyse fliC_fliA fwd	seq. of ΔfliC strains (in a ΔfliA context)	TTGCTCGTGTAGATGATTCC
MZ_107	Tet-SacB_Kassettenerweiterung fwd	amplifies tetA-sacB or fragments for SOEing PCR	CACATGGAAGTTGGAAGTCC TCCTAATTTTTGTTGACACTCTATC
MZ_108	Tet-SacB deltafliC fwd_neu	genomic deletion of fliC	ATCAGGCAATTTGGCGTTGCCGTCAGTCTC AGTTAATCAGGTTACAACGACACATGGAAGTTGGAAGTCC
MZ_109	TetA-sacB deltacspD_1	genomic deletion of cspD	CGATCGGGCTGGCATTTTGCCTTTAGGATGTACACAA TGAGACAGAAGAGCACATGGAAGTTGGAAGTCC
MZ_110	TetA-sacB deltacspD_2	genomic deletion of cspD	CCCGGTTTATTCCATCTTACTTGTATAAGATTTGCG AAGGATGTCGAAGCATCAAAGGGAAAACTGTCCATATGC
MZ_111	deltacspD_TS1f	genomic deletion of cspD	GCAGTAAAGTGCTGCGTG
MZ_112	deltacspD_TS1r	genomic deletion of cspD	ATGTCGAAGCCTCTTCTGTCTCATTGTGTACATC
MZ_113	deltacspD_TS2f	genomic deletion of cspD	GACAGAAGAGGCTTCGACATCCTTCGC
MZ_114	deltacspD_TS2r	genomic deletion of cspD	CTTGTTCACCATCGCCACTT
MZ_115	Deletionsanalyse cspD fwd	seq. of ΔcspD strains	AACAGTCGATGTTTGGTAGC
MZ_116	Deletionsanalyse cspD rev	seq. of ΔcspD strains	AGCAATGGATGCTCATTCTC
MZ_119	deltaflg_TS1f	genomic deletion from flgN to flgL	GCGTTGGGCATTCTTTCC
MZ_120	deltaflg_TS1r	genomic deletion from flgN to flgL	AGATTATCTCCGGCCTGCAC
MZ_121	deltaflg_TS2f	genomic deletion from flgN to flgL	GTGCAGGCCGGAGATAATCTTTTC GCTTTAAAACATATCATGAA

No.	Name	Function	Sequence 5′ -> 3′
MZ_122	deltaflg_TS2r	genomic deletion from flgN to flgL	GTCTGATGTTGCCGTAGC
MZ_123	Deletionsanalyse flg fwd	seq. of ΔflgNMABCDEFGHIJKL strains	GTCTTGGATGTATTACGCCG
MZ_124	Deletionsanalyse flg rev	seq. of ΔflgNMABCDEFGHIJKL strains	CAAAGTCTGGATCCGCTATC
MZ_125	Deletionsanalyse flg WT rev	seq. of ΔflgNMABCDEFGHIJKL strains (WT var)	CTGGCGACGCTGGATTA
MZ_132	TetA-sacB deltaflg_1_neu	genomic deletion from flgN to flgL	GGACGGTGTAACAATGCATTCCGGCCTGCA GTGCAGGCCGGAGATAATCTCACATGGAAGTTGGAAGTCC
MZ_133	TetA-sacB deltaflg_2_neu	genomic deletion from flgN to flgL	GAGCAGGCAGACAAAACATACCCAGTTTCATGATA TGTTTTAAAGCGAAAATCAAAGGGAAAACTGTCCATATGC
MZ_134	TetA-sacB deltafliE_1	genomic deletion from fliE to fliR	CCATGCCCCACGCGGTGAGATCGGATGTAATTAA CAACAGGTTACAAAACCACATGGAAGTTGGAAGTCC
MZ_135	TetA-sacB deltafliR_2	genomic deletion from fliE to fliR	GGTATTATTTTTCGGATAATCCTTAGGATAACATGA TAAACGTTACGGAAATCAAAGGGAAAACTGTCCATATGC
MZ_136	TetA-sacB deltaflh_1	genomic deletion from flhE to motA	TTTTCACTGAGTTATTAAACATACTCGCGAGCGCGTAA TTTTTTTGTCCTCACATGGAAGTTGGAAGTCC

Supporting information S2: Experimental procedures (additional information)

Construction of Deletion Strains

For each deletion locus, two cycles of recombination were used to first insert the dual-selectable *tetA-sacB* cassette and second to replace it with a fused DNA construct of sequences adjacent to the deletion locus. Successful recombination with the construct effectively resulted in the deletion of the target locus. Initially, the original strain was transformed with pSIM5. pSIM5 contains the *exo*, *bet*, and *gam* genes of bacteriophage lambda under transcriptional control of a temperature inducible system allowing induction at 42 °C and normal growth at 30 °C. Moreover, pSIM5 has a temperature-sensitive origin of replication that enables curing of the plasmid upon cultivation in non-selective media at 37 °C.

For the first recombineering cycle, the *tetA-sacB* cassette was amplified from the genome of T-SACK and used as a template for amplifying DNA fragments carrying about 50 bp of homologous arms for chromosomal recombination. If necessary, homologous arms were added separately to *tetA* and *sacB* and the fragments joined by overlap PCR. A strain carrying pSIM5 was then cultivated in 10 ml 2xTY medium in a 50 ml baffled shaking flask at 30 °C on a rotary shaker set to 130 rpm until an OD of 0.2 was reached. The culture was transferred into a 42 °C water bath and agitated for 15 min to induce recombineering proteins from pSIM5. Cells were then chilled on ice and electro-competent cells were prepared following standard protocols. Cells were transformed with about 100 ng of the *tetA-sacB* PCR product carrying homologous arms and regenerated in SOC medium at 37 °C for 3 - 4 hours. Regenerated cells were plated on 2xTY agar plates containing tetracycline. A single colony was picked and again transformed with pSIM5 to yield a recombineering competent intermediate strain carrying a chromosomal copy of the *tetA-sacB* cassette in the target locus.

For the second cycle of recombination, first the recombineering template was constructed by amplification of DNA sequences from the genome of *E. coli* MG1655 adjacent to the deletion locus and joining of these sequences by overlap PCR to create a fused construct of the neighboring regions. The homologous arms created in this step ranged from about 200 to 500 bp. The intermediate strain was subjected to the same procedure as described for the first recombination cycle. Cells were then transformed with about 100 ng of the target DNA sequence, regenerated in SOC medium at 37 °C for 3 - 4 hours and

plated on counterselection agar. After incubation for 2 days at 42 °C large colonies were picked and tested for the absence of tetracycline resistance by streaking onto 2xTY agar plates containing tetracycline. Colonies showing no growth on tetracycline agar plates were cultivated in 2xTY at 37 °C, their DNA isolated and the presence of the deletion locus verified by sequencing.

Construction of eGFP production strains

Recombineering of *E. coli* MG1655 and *E. coli RM214* was conducted as described in the previous section. For the first cycle of recombineering a *tetA-sacB* cassette with homology to the *rhaB* gene was created by amplifying the *tetA-sacB* cassette from the genome of T-SACK with primers 427 and 428. The resulting intermediate strain was then subjected to a second cycle of recombineering with a copy of the inactive *rhaB* gene amplified from the genome of *E. coli* BW3110 pJOE4056.2 with primers 425 and 426. After counter-selection and testing for the absence of tetracycline resistance the presence of the frameshift mutation was verified by sequencing and the inability of the strains to ferment rhamnose confirmed by streaking on 2xTY indicator agar plates containing rhamnose.

The backbone of pJOE4056.2 was amplified with primers 432 and 433. The *tetA* cassette was amplified from the genome of T-SACK with primers 107 and 431. The two fragments were joined and circularized in a Gibson assembly reaction (Gibson et al., 2009) to yield pJOE4056.2_tetA and *E. coli* DH5α λpir was transformed with the reaction product by electroporation. After verification of the plasmid by partial sequencing, *E. coli* MG1655 *rhaB*$^-$ and *E. coli* RM214 *rhaB*$^-$ were transformed with pJOE4056.2_tetA. Finally, the resulting strains were streaked on 2xTY indicator plates with rhamnose to confirm both the absence of rhamnose catabolism as well as the induction of eGFP production in presence of rhamnose and absence of glucose.

Purification and Quantification of eGFP

E. coli BW3110 pJOE4056.2 was cultivated in glucose-limited minimal medium in preliminary experiments. Cells were harvested by centrifugation, resuspended in homogenization buffer (14 mM magnesium acetate, 60 mM potassium acetate, 10 mM Tris pH 8.0, 2 mM dithiotreitol) using approximately 1 ml buffer per 1 g of cell mass and passed twice through a high pressure homogenizer at 15 – 20 kPa. Homogenized cell suspension was

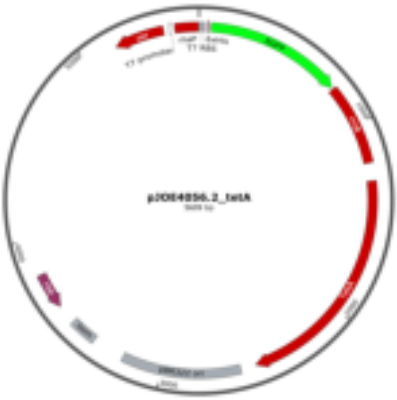


FIGURE B.1: Plasmid map of pJOE4056.2_tetA.

cleared by centrifugation at 15.000 g for 15 min. Purification of his-tagged eGFP was conducted using nickel affinity chromatography (running buffer: 50 mM Tris-HCl pH 7.0, elution buffer: 50 mM tris-HCl pH 7.6, 15 mM Imidazol).

The concentration of purified eGFP stock solution was determined to be 3.236 g/l by Bradford Assay. The stock solution was diluted with ice-cold PBS-MgCa to measure an eGFP calibration curve. 200 µl of diluted stock were transferred into a black 96 well-plate with transparent bottom and lid and the fluorescence (excitation 485 nm, emission 535 nm) quantified in a SLT SpectraFluor plate-reader (Tecan, Switzerland). The resulting calibration curve was used to convert the fluorescent values of bioprocess samples to eGFP concentrations [g/l]:

$$c_{eGFP} = RLU_{sample} * 9.9135 * 10^{-6} + 9.6462 * 10^{-6} \tag{B.1}$$

Bioreactor Setup

The primary reactor was a 3 L bioreactor (Bioengineering, Switzerland) equipped with flow baffles and two six-blade Rushton type impellers operated at 1000 rpm. A constant aeration rate of 2.0 standard liters of ambient pressurized air per minute was employed and the system operated at a total pressure of 1.5 bar. Temperature was monitored by a platinum resistance thermometer and regulated by electrical heating or water cooling. Temperature was set to 37 °C. The reactor was equipped with a pH sensor (Mettler Toledo,

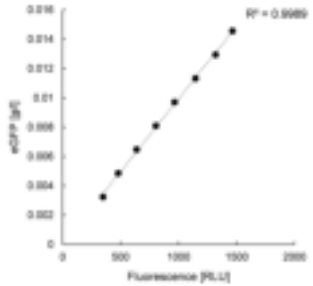


FIGURE B.2: Calibration curve for the conversion of fluorescence units to eGFP concentration.

Columbus, USA) to control pH and a pO2 sensor for monitoring dissolved oxygen tension (PreSens, Regensburg, Germany). During all fermentation stages pH was set to 7.0 and regulated by automated addition of 3 M NaOH or 2.5 M H_3PO_4. Dissolved oxygen tension was not regulated but maintained values above 50% saturation to 1.5 bar ambient air at all times. In the exhaust gas stream, the concentration of oxygen and carbon dioxide was measured by gas sensors (BlueSens, Herten, Germany). During the chemostat phase the feed was constantly added to the reactor by a peristaltic pump (Watson-Marlow, Falmouth, United Kingdom). The feed flow was monitored by a balance recording the weight of the stirred feed barrel and manually adjusted if necessary. The harvesting pump operated as a slave pump set to maintain a constant weight of the bioreactor. For this purpose, the stirred tank reactor was installed on a balance as well.

The secondary compartment was a plug-flow reactor with an inner tube diameter of 20 mm and a total volume of approximately 380 ml. Five ports along the primary axis were used to take samples throughout the cultivation. Next to the first port P1 additional constant aeration was provided at 0.15 standard liters per minute. In the PFR, oxygen saturation was monitored close to ports P1 and P5 and maintained levels above 30% saturation of ambient air conditions throughout the cultivation. Temperature in the PFR was maintained at 36 – 37 °C by water heating and isolation material. A diaphragm metering pump (Sigma/1, ProMinent, Heidelberg, Germany) was used to transfer biosuspension from the stirred tank reactor the plug flow reactor after connection of the two reactors.

Sequence analysis of *E. coli* RM214

Conducting the genomic deletions required a high number of total passages until *E. coli* RM214 was completed. We sequenced both the genome of *E. coli* MG1655 and *E. coli* RM214 to ensure no detrimental mutations or rearrangements had occurred (see also Supplementary Information S8). We first compared the sequence of our *E. coli* MG1655 isolate to the reference sequence NC_000913.3 which revealed the presence of two known sequence variations in different isolates of *E. coli* MG1655 affecting *gatC* and *glpR* (Freddolino et al., 2012). We found additional SNPs in *gfcD*, *yciG*, *wbbI* and an insertion in an intergenic region, none of which appear to confer a detrimental phenotype in standard cultivations. Sequence analysis of *E. coli* RM214 revealed additional SNPs in *elfC*, *trmD*, *dcuD*, the reversion of the SNP in *gfcD* and multiple SNPs in *insH5* in the rac prophage region. All these mutations are irrelevant as *elfC* is only involved in pathogenicity, *dcuD* encodes a weakly expressed C4-carboxylate transporter not beneficial in glucose minimal medium and *insH5* is known to be a mutational hotspot. TrmD is an essential tRNA-methyltransferase and the mutation confers an amino acid exchange [H78N]. However, inspection of the structure of TrmD revealed that the exchange had happened far away from the catalytically active center thus presumably not affecting the biological function. Our considerations are supported by the normal growth phenotype exhibited by all strains of the deletion series.

Supporting information S3: Flow cytometry histograms

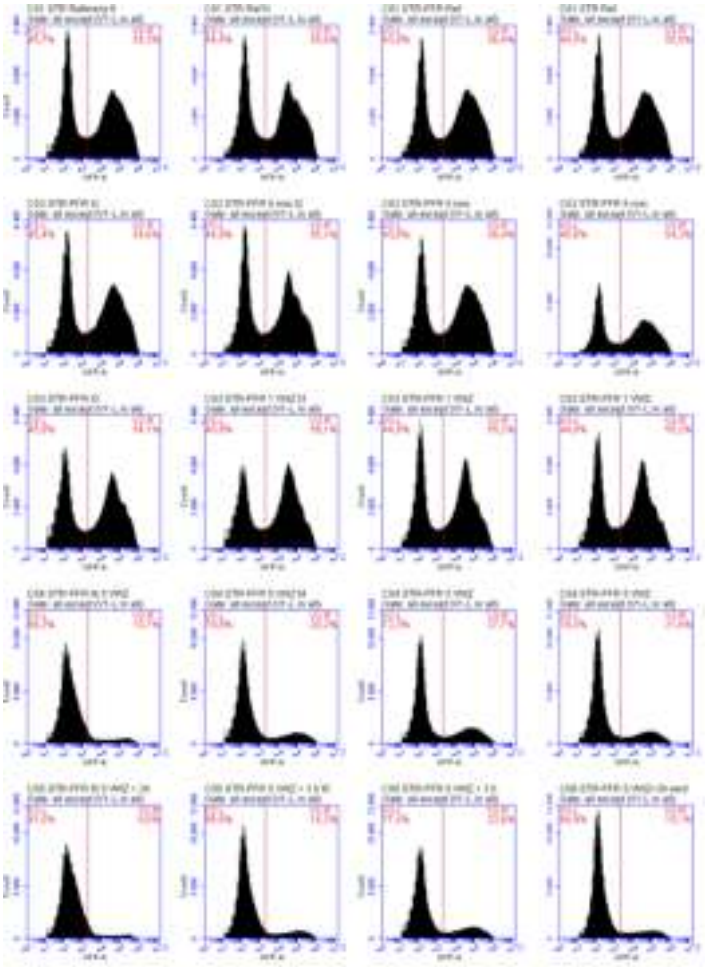


FIGURE B.3: Flow cytometry histograms of green fluorescence signals of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. The vertical red line (2000 RLU in signal GFP-A) indicates the division line used to separate non-producer cells (V1-L) from producer cells (V1-R). Each column of histograms was recorded from a single fermentation (n = 4 for each strain). Rows indicate the time-point of sampling (1st: STR reference sample, 2nd: STR-PFR 5 min, 3rd: STR-PFR 5 h, 4th: STR-PFR 25 h, 5th: STR-PFR 28 h).

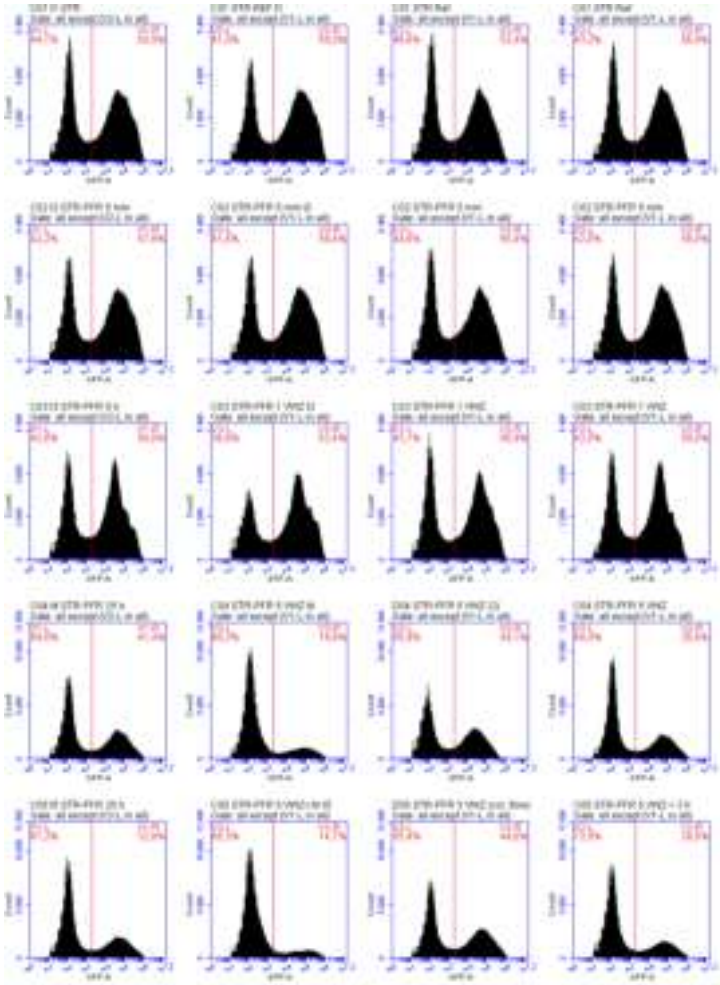


FIGURE B.4: Flow cytometry histograms of green fluorescence signals of *E. coli* RM214 rhaB^{-} pJOE4056.2_tetA. The vertical red line (2000 RLU in signal GFP-A) indicates the division line used to separate non-producer cells (V1-L) from producer cells (V1-R). Each column of histograms was recorded from a single fermentation (n = 4 for each strain). Rows indicate the time-point of sampling (1st: STR reference sample, 2nd: STR-PFR 5 min, 3rd: STR-PFR 5 h, 4th: STR-PFR 25 h, 5th: STR-PFR 28 h).

Supporting information S4: AxP concentrations and statistical evaluation.

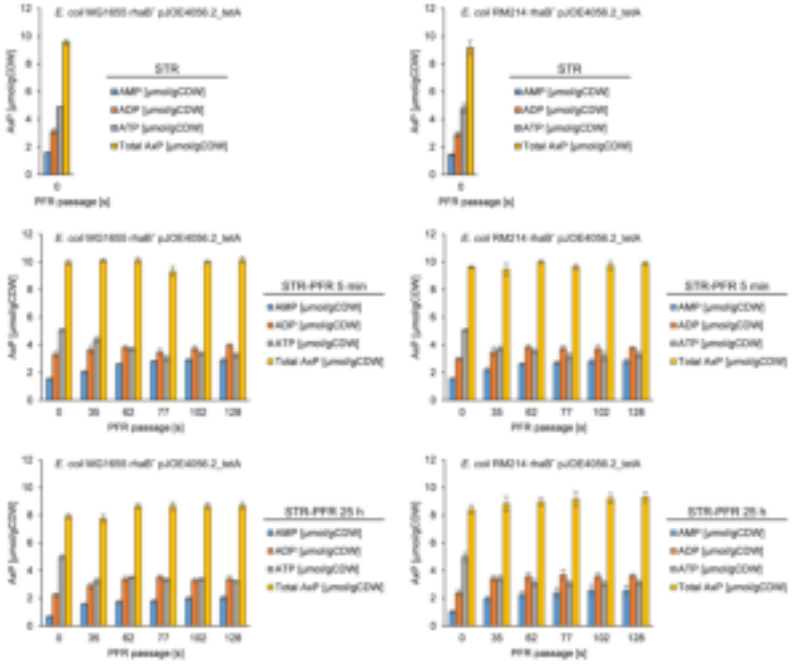


FIGURE B.5: AxP concentrations. Individual cellular levels of AMP, ADP, ATP and the total of all three substances along the PFR passage. The STR sample is indicated at 0 s, the following five samples correspond to the five ports along the primary axis of the PFR and are indicated at the respective mean residence time. The figures show samples from different time-points of the cultivations as given in the figure legend (STR, STR-PFR 5 min, STR-PFR 25 h). **Left:** *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. **Right:** *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. Error bars represent SEM (n = 4).

TABLE B.2: Statistical evaluation of AEC values from the primary reactor (STR)

intra-strain comparison									inter-strain comparison	
	E. coli MG1655 rhaB$^-$ pJOE4045.2_tetA				*E. coli* RM214 rhaB$^-$ pJOE4045.2_tetA					
Time point	Mean AEC	SEM	p		Mean AEC	SEM	p		p	
STR	0.673	0.0032	-	n. s.	0.684	0.0056	-	n. s.	0.15	n. s.
STR-PFR 5 min	0.677	0.0038	0.46	n. s.	0.681	0.0060	0.77	n. s.	0.57	n. s.
STR-PFR 25 h	0.774	0.0080	$2.4*10^{-5}$	***	0.734	0.017	0.028	*	0.077	n. s.

Statistics were calculated from n = 4 biological replicates. Significance indicators (two-tailed t-test): *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. For the intra-strain comparison mean AEC values of a single strain at STR-PFR 5 min or STR-PFR 25 h were compared to its STR value. For the inter-strain comparison mean AEC values of the two strains at a single time point were compared.

Supplementary information S5: Carbon Balance

TABLE B.3: Carbon Balance of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA.

Process stage	**CDW [%]**	**Acetate [%]**	**Rhamnose [%]**	**CO_2(g) [%]**	**CO_2(aq) [%]**	**Total [%]**
STR	36.3 ± 0.8[a]	0.5 ± 0.1	7.9 ± 0.1	49.9 ± 0.8	3.1 ± 1.0	97.7 ± 2.4
STR-PFR 5 min	34.6 ± 0.6	0.4 ± 0.1	7.9 ± 0.1	53.2 ± 0.5	1.5 ± 0.6	97.7 ± 0.8
STR-PFR 5 h	35.8 ± 0.3	0.7 ± 0.1	7.9 ± 0.1	47.9 ± 0.8	1.7 ± 0.4	94.0 ± 0.9
STR-PFR 25 h	38.1 ± 0.7	0.2 ± 0.1	7.9 ± 0.1	48.1 ± 0.3	1.8 ± 0.5	96.0 ± 1.0
STR-PFR 28 h	38.7 ± 0.5	0.1 ± 0.1	7.9 ± 0.1	47.9 ± 0.4	1.8 ± 0.4	96.3 ± 0.6

[a]: Errors indicate SEM (n = 4).

TABLE B.4: Carbon balance of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA

Process stage	**CDW [%]**	**Acetate [%]**	**Rhamnose [%]**	**CO_2(g) [%]**	**CO_2(aq) [%]**	**Total [%]**
STR	37.3 ± 0.6	0.3 ± 0.1	7.9 ± 0.1	48.9 ± 0.3	3.5 ± 0.7	97.9 ± 1.1
STR-PFR 5 min	35.5 ± 0.7	0.3 ± 0.1	7.9 ± 0.1	52.5 ± 1.1	3.4 ± 0.7	99.6 ± 2.0
STR-PFR 5 h	36.2 ± 0.3	0.7 ± 0.2	7.9 ± 0.1	48.0 ± 1.0	3.3 ± 0.6	96.1 ± 1.4
STR-PFR 25 h	37.8 ± 0.7	0.3 ± 0.1	7.9 ± 0.1	47.3 ± 0.3	3.3 ± 0.7	96.7 ± 0.9
STR-PFR 28 h	38.3 ± 0.5	0.2 ± 0.1	7.9 ± 0.1	47.1 ± 0.2	3.6 ± 0.8	97.1 ± 0.7

[a]: Errors indicate SEM (n = 4).

Supplementary information S6: Exhaust Gas Parameters

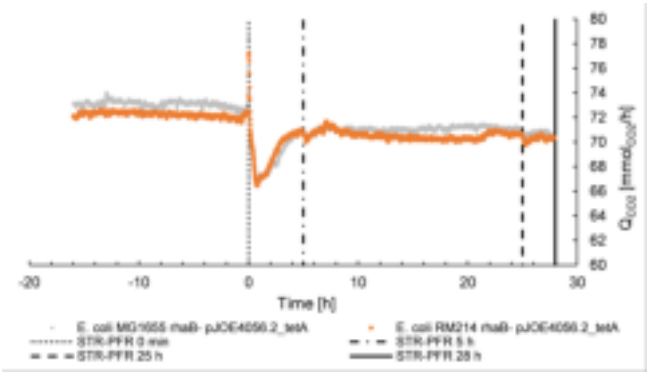


FIGURE B.6: Exhaust Gas Parameters. Oxygen uptake rate QO2. Values represent means of four fermentations with data synchronized to t = 0 h. Technical measurement artifacts were eliminated prior to data analysis. Vertical lines indicate relevant process timepoints: PFR connection (STR-PFR 0 min) and sampling points after 1 volumetric residence time (STR-PFR 5 h), after 5 volumetric residence times (STR-PFR 25 h) and prior to abortion of the fermentation (STR-PFR 28 h).

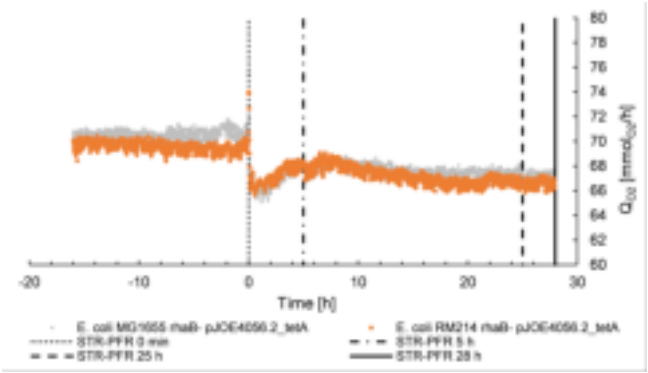


FIGURE B.7: Exhaust Gas Parameters. Carbon Dioxide formation rate QCO2. Values represent means of four fermentations with data synchronized to t = 0 h. Technical measurement artifacts were eliminated prior to data analysis. Vertical lines indicate relevant process timepoints: PFR connection (STR-PFR 0 min) and sampling points after 1 volumetric residence time (STR-PFR 5 h), after 5 volumetric residence times (STR-PFR 25 h) and prior to abortion of the fermentation (STR-PFR 28 h).

Supplementary information S7: Cell Dry Weight and eGFP production

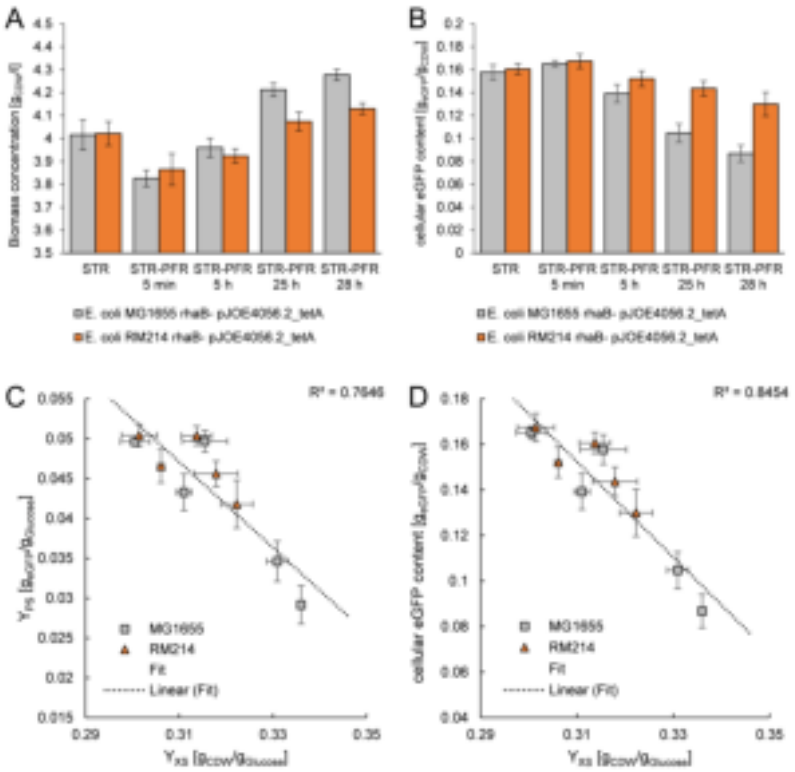


FIGURE B.8: Cell Dry Weight and eGFP production. A: Biomass concentration. B: Cellular eGFP content. C: Correlation of declining product yield and increasing biomass yield. D: Correlation of declining cellular eGFP content and increasing biomass yield. Correlations were calculated from all data points. Error bars represent SEM (n = 4).

Supplementary Data S8: Estimation of ATP demand for eGFP production

The maintenance model by Pirt describes the partition of substrate uptake for growth or maintenance demand (Pirt, 1965).

$$q_s = q_{s,\mu} + m_S \tag{B.2}$$

In analogy, we can differentiate between ATP demand for growth and ATP demand for maintenance.

$$q_{ATP} = q_{ATP,X} + m_{ATP} \tag{B.3}$$

Assuming complete oxidation of substrate used for maintenance requirements and an ATP formation of $Y_{ATP,S}$ = 16.4 $mmol_{ATP}$ $mmol^{-1}_{Glucose}$ (2 mol ATP from glycolysis, P/O-Ratio of 1.2 in the respiratory chain), we can estimate m_{ATP} from m_S.

$$m_{ATP} = m_S * Y_{ATP,S} \tag{B.4}$$

Inserting data from the heterogeneous STR-PFR conditions yields:
E. coli MG1655: $m_{ATP} = m_S * Y_{ATP,S} = 0.14\ \frac{g_{Glucose}}{g_{CDW}*h} * 16.4 \frac{mol_{ATP}}{mol_{Glucose}*h} = 12.74 \frac{mmol_{ATP}}{g_{CDW}*h}$
E. coli RM214: $m_{ATP} = m_S * Y_{ATP,S} = 0.10\ \frac{g_{Glucose}}{g_{CDW}*h} * 16.4 \frac{mol_{ATP}}{mol_{Glucose}*h} = 9.10 \frac{mmol_{ATP}}{g_{CDW}*h}$
With the same assumptions as before (2 mol ATP from glycolysis, P/O-Ratio of 1.2 in the respiratory chain), we can calculate the total ATP generation of *E. coli* MG1655 at μ = 0.2 h^{-1} from measured glucose and oxygen uptake rates during the maintenance process at D = 0.2 h^{-1} at the STR-PFR 25 h time-point (data not shown). This yields an ATP generation of q_{ATP} = 39.39 $mmol_{ATP} g^{-1}_{CDW} h^{-1}$. In balanced growth ATP generation equals ATP consumption.
Using equation (B.3) we can now calculate the ATP demand of *E. coli* MG1655 for growth at μ = 0.2 h^{-1}:

$$q_{ATP,X} = q_{ATP} - m_{ATP} = 26.65\ \frac{mmol_{ATP}}{g_{CDW}h} \tag{B.5}$$

Since $Y_{XS,true}$ is not significantly different, we know that the true biomass yield for both strains is identical. Since the metabolic pathways are identical in both strains, the ATP demand for growth $q_{ATP,X}$ can then be assumed to be identical for both strains as well.

In order to calculate the ATP demand for eGFP production we need to introduce a parameter describing ATP demand for eGFP synthesis into equation (B.3). This parameter, $q_{ATP,eGFP}$, includes any additional cellular ATP expenses due to heterologous protein

expression:

$$q_{ATP} = q_{ATP,X} + m_{ATP} + q_{ATP,eGFP} \tag{B.6}$$

We can calculate the total ATP generation of both production strains at μ = 0.2 h^{-1} from measured glucose and oxygen uptake rates during the eGFP production processes at time-point STR-PFR 25 h (data not shown). This estimation yields an ATP generation of q_{ATP} = 45.37 $mmol_{ATP}$*g_{CDW}^{-1}*h^{-1} for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and q_{ATP} = 45.95 $mmol_{ATP}$*$gCDW^{-1}$*h^{-1} for *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA.

Using these values for total ATP generation and the estimated values of the base strains for m_{ATP} and $q_{ATP,X}$ at μ = 0.2 h^{-1} from the maintenance processes, we can finally estimate the ATP demand of eGFP formation by inserting all values into equation (B.6):

$$q_{ATP,eGFP} = q_{ATP} - q_{ATP,X} - m_{ATP} \tag{B.7}$$

The results are documented in table B.5, and the proportion of ATP used for eGFP production is given.

TABLE B.5: Estimated ATP demand of eGFP synthesis at STR-PFR 25 h

Strain	$\mathbf{q}_{ATP,eGFP}$ **[mmol$_{ATP}$/(g$_{CDW}$h)]**	$\mathbf{q}_{ATP,eGFP}$/$\mathbf{q}_{ATP}$ **[%]**
***E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA**	6.0 ± 0.59[a]	13 ± 1.3
***E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA**	10.2 ± 0.58	22 ± 1.7

[a]: Errors indicate SEM (n = 4).

Supporting information S9: SNPs found after continuous cultivation

Each table in this section contains the results of a genomic SNP analysis performed. Cryo MG1655 VS NC_000913.3 compares our isolate (cryostock) of MG1655 to the refseq. Cryo RM214 VS Cryo MG1655 shows SNPs accumulated during strain construction of RM214.

The remaining tables compare sequenced samples from the end of continuous chemostat cultivations to the respective cryostock sample.The numbers indicate the dilution rate used during the chemostat phases (D005 means D = 0.05 h^{-1} and so on).

TABLE B.6: Cryo MG1655 VS NC_000913.3

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
NC_000913.3	1047882	snp	C	A	A:336 C:0	gfcD	Irrelevant, gfcD is not transcribed in E. coli K-12.
NC_000913.3	1315865	snp	G	T	T:300 G:0	yciG	Silent mutation
NC_000913.3	2105253	snp	T	A	A:310 T:0	wbbI	[I269F] of WbbI (beta-1,6-galactofuranosyl-transferase)
NC_000913.3	2173360	del	ACC	A	A:225 ACC:0	gatC	Sequence variation in different MG1655 isolates has been documented previously.
NC_000913.3	3560455	ins	C	CG	CG:335 C:0	glpR	Sequence variation in different MG1655 isolates has been documented previously.
NC_000913.3	4296380	ins	A	ACG	ACG:169 A:0		Intergenic Repeat Region

TABLE B.7: Cryo RM214 VS Cryo MG1655

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4504256-4522705	290658	snp	G	C	C:11 G:0		Low Coverage
2:4504256-4522705	566050	snp	N	T	T:30 N:0		Assembly Gap
2:4504256-4522705	998716	snp	T	G	G:325 T:0	elfC	V -> G at amino acid position 67 of ElfC. Cryptic operon that is involved in pathogenicity.
2:4504256-4522705	1047183	snp	A	C	C:306 A:0	gfcD	Identical to NC_000913.3 MG1655 Refseq, potentially a true polymorphism or reversion.
2:4504256-4522705	1128700	snp	G	T	T:51 G:0		Intergenic region downstream of murJ
2:4504256-4522705	1392292	snp	T	C	C:314 T:0		Intergenic region between pgrR and mppa
2:4504256-4522705	1396700	mnp	NNN	ACA	ACA:14 NNN:0		Assembly Gap
2:4504256-4522705	1427166	snp	A	G	G:118 A:0	ins-H5	ins-H5, rac Prophage
2:4504256-4522705	1427200	snp	T	C	C:55 T:0		
2:4504256-4522705	1427218	snp	T	C	C:33 T:0		
2:4504256-4522705	1427749	snp	G	A	A:55 G:1		

TABLE B.7: Cryo RM214 VS Cryo MG1655

2:4504256-4522705	1427757 snp	A	G	G:83 A:1		
2:4504256-4522705	1427776 snp	C	T	T:115 C:1		
2:4504256-4522705	1427788 snp	A	G	G:146 A:1		
2:4504256-4522705	1427812 snp	A	G	G:189 A:0		
2:4504256-4522705	1427824 complex	CGCG	GGCA	GGCA:210 CGCG:0		
2:4504256-4522705	1527658 complex	NN	AA	AA:16 NN:0		Assembly Gap
2:4504256-4522705	1978629 snp	N	T	T:18 N:0		Assembly Gap
2:4504256-4522705	2304179 mnp	NNN	AAC	AAC:19 NNN:0		Assembly Gap
2:4504256-4522705	2725454 complex	NN	AT	AT:11 NN:0		Assembly Gap
2:4504256-4522705	2744523 snp	G	T	T:295 G:0	trmD	[H78N] of TrmD (tRNA-methyltransferase), TrmD is essential, the exchange is located in an alpha-helix on the surface of the protein and likely does not interfere with the active center [see: Elkins et al. (2003)]

TABLE B.7: Cryo RM214 VS Cryo MG1655

2:4504256-4522705	3375243	snp	T	C	C:277 T:0	dcuD	[V321A] of DcuD (C4-dicarboxylate transporter), dcuD is weakly transcribed and may be involved in glycerol metabolism.

TABLE B.8: MG1655 D005 VS MG1655 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4504256-4522705	390972	snp	N	A	A:32 N:0		Assembly Gap
2:4504256-4522705	566050	snp	N	T	T:19 N:0		Assembly Gap
2:4504256-4522705	729093	snp	N	T	T:11 N:0		Assembly Gap
2:4504256-4522705	1396700	mnp	NNN	ACA	ACA:14 NNN:0		Assembly Gap
2:4504256-4522705	1427166	snp	A	G	G:98 A:0	ins-H5	ins-H5, rac Prophage
2:4504256-4522705	1427200	snp	T	C	C:50 T:0		
2:4504256-4522705	1427218	snp	T	C	C:25 T:0		
2:4504256-4522705	1427749	snp	G	A	A:52 G:1		
2:4504256-4522705	1427757	snp	A	G	G:65 A:1		
2:4504256-4522705	1427776	snp	C	T	T:100 C:0		
2:4504256-4522705	1427788	snp	A	G	G:121 A:0		

TABLE B.8: MG1655 D005 VS MG1655 Cryo

2:4504256-4522705	1427812	snp	A	G	G:155 A:0		
2:4504256-4522705	1427824	complex	CGCG	GGCA	GGCA:176 CGCG:0		
2:4504256-4522705	1978629	snp	N	T	T:23 N:0		Assembly Gap
2:4504256-4522705	2725454	complex	NN	AT	AT:22 NN:0		Assembly Gap
2:4504256-4522705	3364958	snp	N	C	C:26 N:0		Assembly Gap
2:4504256-4522705	3423073	mnp	NN	TA	TA:14 NN:0		Assembly Gap
2:4504256-4522705	4497653	snp	N	T	T:33 N:0		Assembly Gap
2:4504256-4522705	4661603	snp	T	C	C:11 T:0		Assembly Gap

TABLE B.9: MG1655 D01 VS MG1655 Cryo

CHROM	**POS**	**TYPE**	**REF**	**ALT**	**EVID**	**Gene**	**Comment**
2:4504256-4522705	390972	snp	N	A	A:14 N:0		Assembly Gap
2:4504256-4522705	566050	snp	N	T	T:12 N:0		Assembly Gap
2:4504256-4522705	1427166	snp	A	G	G:32 A:0	ins-H5	ins-H5, rac Prophage
2:4504256-4522705	1427776	snp	C	T	T:30 C:0		
2:4504256-4522705	1427788	snp	A	G	G:46 A:0		
2:4504256-4522705	1427812	snp	A	G	G:68 A:0		

TABLE B.9: MG1655 D01 VS MG1655 Cryo

2:4504256-4522705	1427824 complex	CGCGGGCA	GGCA:80 CGCG:0		
2:4504256-4522705	2725454 snp	N	A	A:10 N:0	Assembly Gap
2:4504256-4522705	2725455 snp	N	T	T:10 N:0	Assembly Gap
2:4504256-4522705	3364958 snp	N	C	C:12 N:0	Assembly Gap
2:4504256-4522705	3423073 mnp	NN	TA	TA:14 NN:0	Assembly Gap
2:4504256-4522705	4497653 snp	N	T	T:16 N:0	Assembly Gap

TABLE B.14: RM214 D02 VS RM214 Cryo

CHROM	**POS**	**TYPE**	**REF**	**ALT**	**EVID**	**Gene**	**Comment**
2:4455130-4470514	15384	del	CA	C	C:18 CA:0		low coverage
2:4455130-4470514	256558	snp	N	C	C:17 N:0		Assembly Gap
2:4455130-4470514	269203	snp	N	A	A:10 N:0		Assembly Gap
2:4455130-4470514	277833	snp	N	C	C:11 N:0		Assembly Gap
2:4455130-4470514	290081	snp	G	C	C:10 G:0		Assembly Gap
2:4455130-4470514	573286	complex	NNN	CTT	CTT:14 NNN:0		Assembly Gap
2:4455130-4470514	728517	snp	N	T	T:10 N:0		Assembly Gap
2:4455130-4470514	921089	snp	N	A	A:16 N:0		Assembly Gap
2:4455130-4470514	1092976	snp	N	T	T:20 N:0		Assembly Gap

TABLE B.14: RM214 D02 VS RM214 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4455130-4470514	1206538	ins	G	GT	GT:173 G:0	ycfk (e14 prophage)	Identical to NC_000913.3 MG1655 Refseq, potentially true polymorphism
2:4455130-4470514	1394788	mnp	NN	GC	GC:21 NN:0		Assembly Gap
2:4455130-4470514	1426598	snp	A	G	G:81 A:1	ins-H5	ins-H5, rac Prophage
2:4455130-4470514	1426632	snp	T	C	C:48 T:0		
2:4455130-4470514	1426650	snp	T	C	C:27 T:0		
2:4455130-4470514	1427181	snp	G	A	A:93 G:0		
2:4455130-4470514	1427189	snp	A	G	G:111 A:0		
2:4455130-4470514	1427208	snp	C	T	T:148 C:0		
2:4455130-4470514	1427220	snp	A	G	G:166 A:0		
2:4455130-4470514	2022495	snp	N	C	C:24 N:0		Assembly Gap
2:4455130-4470514	2051091	snp	N	T	T:17 N:0		Assembly Gap
2:4455130-4470514	2065039	snp	N	T	T:23 N:0		Assembly Gap
2:4455130-4470514	2173885	mnp	NN	TA	TA:11 NN:0		Assembly Gap

TABLE B.14: RM214 D02 VS RM214 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4455130-4470514	2437787	snp	N	C	C:13 N:0		Assembly Gap
2:4455130-4470514	2661199	snp	N	T	T:14 N:0		Assembly Gap
2:4455130-4470514	3364418	mnp	NN	CA	CA:17 NN:0		Assembly Gap
2:4455130-4470514	3650924	snp	N	A	A:19 N:0		Assembly Gap
2:4455130-4470514	3665101	snp	N	G	G:26 N:0		Assembly Gap
2:4455130-4470514	3766292	snp	N	C	C:74 N:0		Assembly Gap
2:4455130-4470514	4497117	snp	N	T	T:24 N:0		Assembly Gap

TABLE B.10: MG1655 D02 VS MG1655 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4504256-4522705	278410	snp	N	C	C:10 N:0		Assembly Gap
2:4504256-4522705	390972	snp	N	A	A:32 N:0		Assembly Gap
2:4504256-4522705	566050	snp	N	T	T:24 N:0		Assembly Gap
2:4504256-4522705	729093	snp	N	T	T:12 N:0		Assembly Gap
2:4504256-4522705	1396700	snp	N	A	A:11 N:0		Assembly Gap
2:4504256-4522705	1396701	snp	N	C	C:11 N:0		Assembly Gap
2:4504256-4522705	1396702	snp	N	A	A:11 N:0		Assembly Gap
2:4504256-4522705	1427166	snp	A	G	G:107 A:0	ins-H5	ins-H5, rac Prophage
2:4504256-4522705	1427200	snp	T	C	C:55 T:0		
2:4504256-4522705	1427218	snp	T	C	C:25 T:0		
2:4504256-4522705	1427749	snp	G	A	A:64 G:0		
2:4504256-4522705	1427757	snp	A	G	G:78 A:0		
2:4504256-4522705	1427776	snp	C	T	T:115 C:0		
2:4504256-4522705	1427788	snp	A	G	G:131 A:0		
2:4504256-4522705	1427812	snp	A	G	G:182 A:0		
2:4504256-4522705	1427824	complex	CGCG	GGCA	GGCA:198 CGCG:0		
2:4504256-4522705	1978629	snp	N	T	T:22 N:0		Assembly Gap
2:4504256-4522705	2304179	mnp	NNN	AAC	AAC:12 NNN:0		Assembly Gap
2:4504256-4522705	2725454	mnp	NN	AT	AT:15 NN:0		Assembly Gap
2:4504256-4522705	3364958	snp	N	C	C:18 N:0		Assembly Gap
2:4504256-4522705	3423073	mnp	NN	TA	TA:22 NN:0		Assembly Gap
2:4504256-4522705	4497653	snp	N	T	T:30 N:0		Assembly Gap

TABLE B.11: MG1655 D03 VS MG1655 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4504256-4522705	390972	snp	N	A	A:27 N:0		Assembly Gap
2:4504256-4522705	566050	snp	N	T	T:13 N:0		Assembly Gap
2:4504256-4522705	1427166	snp	A	G	G:80 A:0	ins-H5	ins-H5, rac Prophage
2:4504256-4522705	1427200	snp	T	C	C:38 T:0		
2:4504256-4522705	1427218	snp	T	C	C:17 T:0		
2:4504256-4522705	1427749	snp	G	A	A:46 G:0		
2:4504256-4522705	1427757	snp	A	G	G:67 A:0		
2:4504256-4522705	1427776	snp	C	T	T:95 C:0		
2:4504256-4522705	1427788	snp	A	G	G:113 A:0		
2:4504256-4522705	1427812	snp	A	G	G:158 A:0		
2:4504256-4522705	1427824	complex	CGCG	GGCA	GGCA:174 CGCG:0		
2:4504256-4522705	1527658	snp	N	A	A:10 N:0		Assembly Gap
2:4504256-4522705	1978629	snp	N	T	T:18 N:0		Assembly Gap
2:4504256-4522705	2304179	mnp	NNN	AAC	AAC:16 NNN:0		Assembly Gap
2:4504256-4522705	2725454	mnp	NN	AT	AT:15 NN:0		Assembly Gap
2:4504256-4522705	3364958	snp	N	C	C:12 N:0		Assembly Gap
2:4504256-4522705	3423073	mnp	NN	TA	TA:18 NN:0		Assembly Gap
2:4504256-4522705	4497653	snp	N	T	T:30 N:0		Assembly Gap

TABLE B.12: RM214 D005 VS RM214 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4455130-4470514	15384	del	CA	C	C:14 CA:0		Low Coverage
2:4455130-4470514	290081	snp	G	C	C:10 G:0		Assembly Gap
2:4455130-4470514	524561	complex	NN	GG	GG:17 NN:0		Assembly Gap
2:4455130-4470514	566734	snp	G	C	C:14 G:0		Assembly Gap
2:4455130-4470514	1206538	ins	G	GT	GT:246 G:2	ycfk (e14 prophage)	Identical to NC_000913.3 MG1655 Refseq, potentially true polymorphism
2:4455130-4470514	1208356	complex	AN	A	A:233 AN:0	stfE (e14 prophage)	Identical to NC_000913.3 MG1655 Refseq, potentially true polymorphism
2:4455130-4470514	1394788	mnp	NN	GC	GC:14 NN:0		Assembly Gap
2:4455130-4470514	1426598	snp	A	G	G:116 A:0	ins-H5	ins-H5, rac Prophage
2:4455130-4470514	1426632	snp	T	C	C:57 T:0		
2:4455130-4470514	1426650	snp	T	C	C:27 T:0		
2:4455130-4470514	1427181	snp	G	A	A:143 G:0		
2:4455130-4470514	1427189	snp	A	G	G:163 A:0		
2:4455130-4470514	1427208	snp	C	T	T:192 C:0		
2:4455130-4470514	1427220	snp	A	G	G:219 A:0		
2:4455130-4470514	2022495	snp	N	C	C:17 N:0		Assembly Gap
2:4455130-4470514	2065039	snp	N	T	T:21 N:0		Assembly Gap
2:4455130-4470514	2173885	mnp	NN	TA	TA:24 NN:0		Assembly Gap
2:4455130-4470514	2437787	snp	N	C	C:16 N:0		Assembly Gap
2:4455130-4470514	3650924	snp	N	A	A:20 N:0		Assembly Gap
2:4455130-4470514	3665101	snp	N	G	G:14 N:0		Assembly Gap

TABLE B.13: RM214 D01 VS RM214 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4455130-4470514	15384	del	CA	C	C:12 CA:0		Low Coverage
2:4455130-4470514	269203	snp	N	A	A:11 N:0		Assembly Gap
2:4455130-4470514	278787	snp	C	A	A:10 C:0		Assembly Gap
2:4455130-4470514	290081	snp	G	C	C:14 G:0		Assembly Gap
2:4455130-4470514	524561	mnp	NN	GG	GG:23 NN:0		Assembly Gap
2:4455130-4470514	573286	mnp	NNN	CTT	CTT:26 NNN:0		Assembly Gap
2:4455130-4470514	921089	snp	N	A	A:22 N:0		Assembly Gap
2:4455130-4470514	1206538	ins	G	GT	GT:201 G:1	ycfk (e14 prophage)	Identical to NC_000913.3 MG1655 Refseq, potentially true polymorphism
2:4455130-4470514	1208356	complex	AN	A	A:214 AN:0	stfE (e14 prophage)	Identical to NC_000913.3 MG1655 Refseq, potentially true polymorphism
2:4455130-4470514	1426598	snp	A	G	G:118 A:0	ins-H5	ins-H5, rac Prophage
2:4455130-4470514	1426632	snp	T	C	C:66 T:0		
2:4455130-4470514	1426650	snp	T	C	C:36 T:0		
2:4455130-4470514	1427181	snp	G	A	A:132 G:1		
2:4455130-4470514	1427189	snp	A	G	G:154 A:1		
2:4455130-4470514	1427208	snp	C	T	T:194 C:1		
2:4455130-4470514	1427220	snp	A	G	G:209 A:1		
2:4455130-4470514	2022495	snp	N	C	C:21 N:0		Assembly Gap
2:4455130-4470514	2173885	mnp	NN	TA	TA:15 NN:0		Assembly Gap
2:4455130-4470514	2437787	snp	N	C	C:17 N:0		Assembly Gap
2:4455130-4470514	3364418	mnp	NN	CA	CA:26 NN:0		Assembly Gap
2:4455130-4470514	3650924	snp	N	A	A:18 N:0		Assembly Gap

TABLE B.15: RM214 D03 VS RM214 Cryo

CHROM	POS	TYPE	REF	ALT	EVID	Gene	Comment
2:4455130-4470514	15384	del	CA	C	C:12 CA:0		Assembly Gap
2:4455130-4470514	269203	snp	N	A	A:10 N:0		Assembly Gap
2:4455130-4470514	1206538	ins	G	GT	GT:124 G:0	ycfk (e14 prophage)	Identical to NC_000913.3 MG1655 Refseq, potentially true polymorphism
2:4455130-4470514	1208356	complex	AN	A	A:146 AN:0	stfE (e14 prophage)	Identical to NC_000913.3 MG1655 Refseq, potentially true polymorphism
2:4455130-4470514	1426598	snp	A	G	G:26 A:0	ins-H5	ins-H5, rac Prophage
2:4455130-4470514	1427181	snp	G	A	A:28 G:0		
2:4455130-4470514	1427189	snp	A	G	G:36 A:0		
2:4455130-4470514	1427208	snp	C	T	T:58 C:0		
2:4455130-4470514	1427220	snp	A	G	G:74 A:0		
2:4455130-4470514	2051091	snp	N	T	T:12 N:0		Assembly Gap
2:4455130-4470514	2065039	snp	N	T	T:12 N:0		Assembly Gap
2:4455130-4470514	2437787	snp	N	C	C:16 N:0		Assembly Gap

Supporting information S10: RT-qPCR measurements

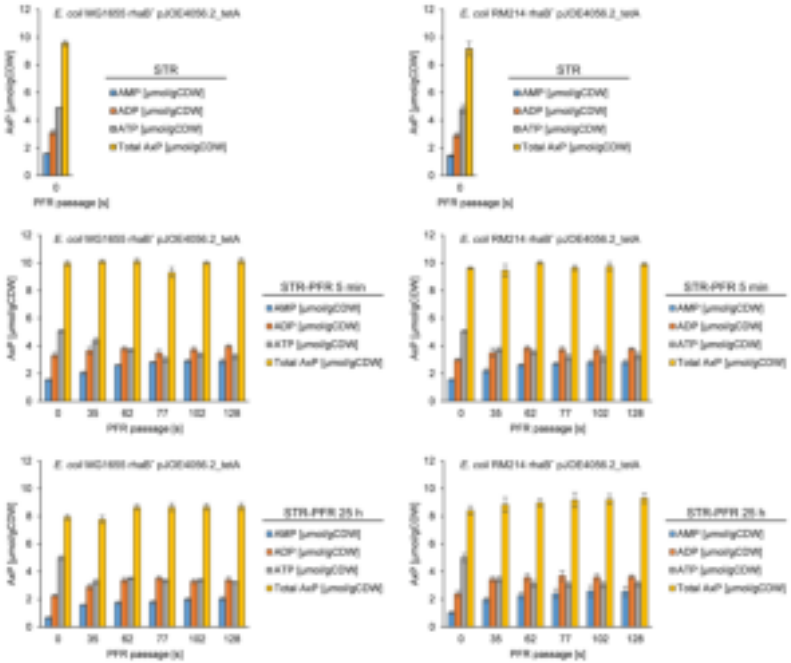


FIGURE B.9: Expression of eGFP. RT qPCR was used to quantify the expression of *egfp* relative to the housekeeping gene *cysG*. The relative expression at time points STR PFR 25 h and STR PFR 28 h was then normalized to the STR sample to yield the fold change in expression compared to the well mixed STR conditions. Mean values from four biological replicates for each strain are shown. Error bars indicate SEM (n = 4).

Appendix C

Transcriptional Profiling of the Stringent Response Mutant Strain *E. coli* SR Reveals Enhanced Robustness to Large-Scale Conditions: Supporting Information

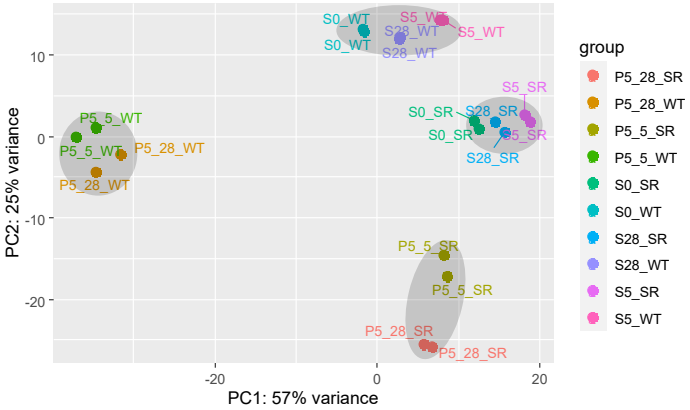
A: Determination of conversion factors from OD 600 nm to cell dry weight concentration

In preliminary experiments, optical density of appropriately diluted fermentation broth from the primary reactor was measured in triplicates at 600 nm on a spectrophotometer (Amersham Biosciences/GE Healthcare, United Kingdom). In parallel, quadruplicate determination of biomass concentration was conducted. 5 ml of broth were centrifuged in weighted glass tubes at 2500 g and 4 °C for 7.5 min. Supernatant was immediately decanted and the pellet washed by resuspending in 5 ml of freshly prepared 150 mM NH_4HCO_3. The suspension was centrifuged again and the washing repeated once. After a final centrifugation the pellet was dried at 105 °C for at least two days and the new weight of the glass tubes measured. The ratio CDW [gL^{-1}] / OD 600 nm was calculated and averaged over all samples to yield conversion factors for each strain. Conversion factors were 0.324 for *E. coli* MG1655 and 0.321 for *E. coli* SR and used in all subsequent experiments to convert OD 600 nm measurements to cell dry weight concentration [gL^{-1}].

TABLE C.1: OD and CDW data from preliminary experiments.

E. coli MG1655			*E. coli SR*		
OD 600 nm	CDW [gL^{-1}]	Ratio CDW / OD	OD 600 nm	CDW [gL^{-1}]	Ratio CDW / OD
Experiment 1			Experiment 1		
8.600	2.545	0.296	8.600	2.975	0.346
8.463	2.875	0.340	8.567	2.925	0.341
8.200	2.550	0.311	8.217	2.815	0.343
8.250	2.555	0.310	8.250	2.865	0.347
Experiment 2			Experiment 2		
8.800	2.720	0.309	9.300	2.665	0.287
8.567	3.140	0.367	9.467	2.790	0.295
8.317	2.925	0.352	9.000	2.715	0.302
8.600	2.685	0.312	9.133	2.790	0.305
Conversion Factor		0.324			0.321

B: Transcriptomic analysis of long term response STR vs STR 0 h



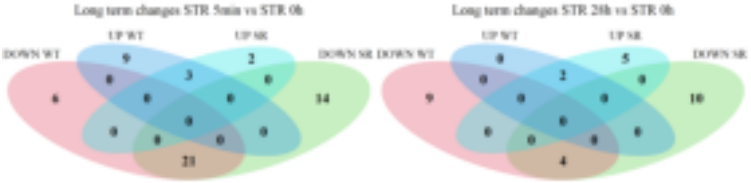


FIGURE C.2: Venn diagrams representing (overlapping) sets of differentially expressed genes derived from repeated ammonia shortage STR-PFR experiments. Long-term response observed for the following comparisons at PFR sample port long term changes (STR vs. STR 0h) conducted after 5 min of process time (left) and 28 h (right). The number of significantly up- and downregulated genes in each set is indicated by numbers. DEGs were defined as having an FDR < 0.01 and $\log_2$fold change > |1|. Complete gene lists of the Venn diagrams are available in the supplementary data table S1 and S2.

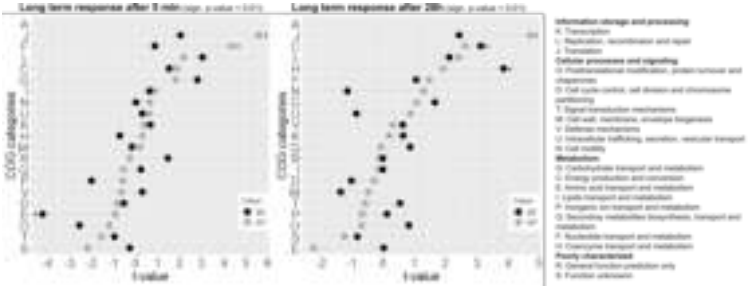


FIGURE C.3: Left: COG categories long term response 5min STR vs STR 0h WT vs SR; Right: COG categories long term response 28h STR vs STR 0h WT vs SR. Sign. categories are indicated with an asterix.

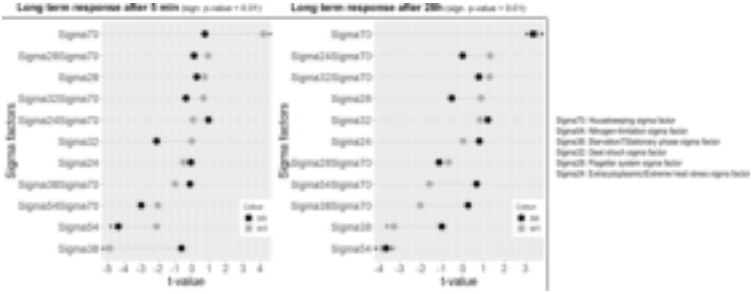


FIGURE C.4: Left: Sigma factors long term response 5 min STR vs STR 0 h WT vs SR. Right: Sigma factors long term response 28 h STR vs STR 0 h WT vs SR. Sign. categories are indicated with an asterix. Sigma 70: Housekeeping; Sigma 54: Nitrogen-limitation; Sigma 38: Starvation/Stationary phase; Sigma 32: Heat shock; Sigma 28: Flagellar system; Sigma 24: Extracytoplasmic/Extreme heat stress

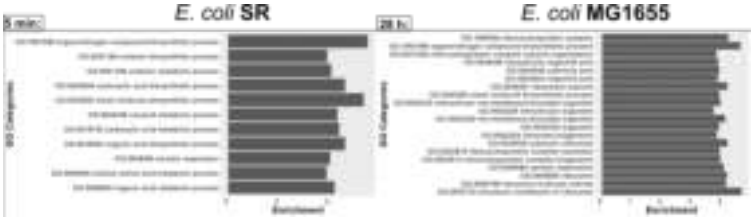


FIGURE C.5: Left: GO categories long term response 28h STR vs STR0 SR. Top 20 significantly upregulated. No sign. down regulated categories. Right: GO categories long term response 28h STR vs STR0 WT. Top 20 significantly upregulated. No sign. down regulated categories.

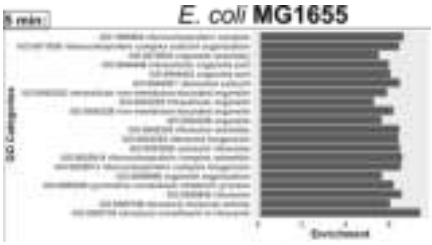


FIGURE C.6: GO categories long term response 5min STR vs STR0 WT. Top 20 significantly upregulated. No sign. down regulated categories.

C: NtrC-mediated operons

TABLE C.2: NtrC-mediated operons (Brown et al., 2014); *E. coli* MG1655

Operons	Log2FC[a]		Function/Pathway
	5min short	28h long	
glnK-amtB	3.81	4.09	GlnK - Nitrogen regulatory protein, AmtB – ammonia transport
gltIJKL	1.14	0.94	Glutamate / aspartate ABC transport
dicC			DNA-binding transcriptional repressor
glnHPQ	3.41	3.37	Glutamine ABC transport
yeaGH	2.98	2.69	YeaG is a serine protein kinase
ycdMLKJIHG	4.01	4.05	Pyrimidine degradation
flgMN	0.29	0.84	Regulation of flagellar synthesis and flagellar biosynthesis protein
ddpXABCDF	4.25	4.49	D-ala-D-ala dipeptide transport and dipeptidase
astCADBE	4.84	4.81	Arginine catabolic patway
fliC	0.46	0.41	Flagellar biosynthesis component
Nac-cbl	4.14	3.94	Nitrogen limitation response-adpater for sigma 70 dependent genes
hisJQMP	1.24	1.64	Histidine ABC transporter
argT	3.32	3.44	Lysine/arginine/ornithine ABC transporter
relA	1.29	1.60	GDP pyrophosphokinase involved in stringent reponse
ssrS	-	-	6S RNA involved stationary phase regulation of transcription
ygjG	3.03	2.79	Putrescine degradative pathway
yhdWXYZ	2.75	2.36	Polar amino acid transport
glnALG	2.57	1.99	Glutamine biosynthesis pathway (ammonia assimilation) and nitrogen regulation
soxR	0.30	0.27	SoxR transcriptional regulator
yjcZ-proP	-0.39	-0.26	Hypothetical protein (YjcZ) + symporter (proP)
potFGHI	0.92	0.82	Putrescine transport

[a] Underlining indicates significant differential expression. Logarithmic ratios are always given for the first gene in the transcription unit.

TABLE C.3: NtrC-mediated operons (Brown et al., 2014); *E. coli* SR

Operons	Log2FC[a]		Function/Pathway
	5min short	28h long	
glnK-amtB	4.05	2.67	GlnK - Nitrogen regulatory protein, AmtB – ammonia transport
gltIJKL	0.25	-0.29	Glutamate / aspartate ABC transport
dicC	-0.09	-0.25	DNA-binding transcriptional repressor
glnHPQ	2.82	2.45	Glutamine ABC transport
yeaGH	1.30	0.89	YeaG is a serine protein kinase
ycdMLKJIHG	3.18	2.83	Pyrimidine degradation
flgMN	0.63	0.73	Regulation of flagellar synthesis and flagellar biosynthesis protein
ddpXABCDF	4.71	3.12	D-ala-D-ala dipeptide transport and dipeptidase
astCADBE	4.11	3.93	Arginine catabolic patway
fliC	0.05	0.17	Flagellar biosynthesis component
Nac-cbl	3.13	1.42	Nitrogen limitation response-adpater for sigma 70 dependent genes
hisJQMP	0.77	0.59	Histidine ABC transporter
argT	2.51	2.01	Lysine/arginine/ornithine ABC transporter
relA	1.35	1.79	GDP pyrophosphokinase involved in stringent reponse
ssrS	-	-	6S RNA involved stationary phase regulation of transcription
ygjG	2.06	2.16	Putrescine degradative pathway
yhdWXYZ	2.00	1.21	Polar amino acid transport
glnALG	2.64	-0.11	Glutamine biosynthesis pathway (ammonia assimilation) and nitrogen regulation
soxR	0.34	1.12	SoxR transcriptional regualtor
yjcZ-proP	-0.77	-0.39	Hypothetical protein (YjcZ) + symporter (proP)
potFGHI	0.44	-1.91	Putrescine transport

[a] Underlining indicates significant differential expression. Logarithmic ratios are always given for the first gene in the transcription unit.

D: Motility Assay of *E. coli* MG1655

A 5 ml overnight preculture of *E. coli* MG1655 in 2xTY medium was incubated at 37 °C and 130 rpm. On the next day, a tube containing motility agar (10 g/l tryptone, 3 g/l yeast extract, 5 g/l NaCl, 0.5 g/l triphenyl-tetrazolium chloride and 0.4 % agar-agar) was inoculated by dipping a thin sterile steel wire into the liquid preculture and stabbing the wire into the motility agar. The culture was incubated without shaking at 37 °C overnight. The spreading of cells as indicated by the expanse of red color from reduction of triphenyl-tetrazolium chloride to triphenylformazan was recorded on the next morning. The formation of a cloudy deep red colored zone indicates strong motility and cell viability.

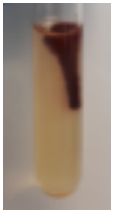


FIGURE C.7: Motility Assay of *E. coli* MG1655.

E: Glucose-specific phospho-transferase system components

TABLE C.4: Logarithmic expression ratio and percentage change of mean expression levels from three glucosespecific phosphotransferase system components.

Gene	Log2FC[a] of *E. coli* SR vs MG1655 (S/S)		Relative Mean Expression Level *E. coli* SR to MG1655	
	5min	28h	5min	28h
ptsG	-0.35	0.67	+ 21.58 %	+ 58.98 %
ptsH	0.57	0.68	+ 48.01 %	+ 60.32 %
ptsI	0.31	0.23	+ 23.82 %	+ 17.28 %

[a] Underlining indicates significant differential expression.

Appendix D

Data-Driven *In-silico* Prediction of Regulation Heterogeneity and ATP Demands of *Escherichia coli* in Large-scale Bioreactors: Supporting Information

D.1 Supporting information A:

The relative distribution of specific mRNAs is taken from the measured normalized counts in the form of transcripts per million. The molecular weight of different mRNAs was calculated with equation 6.1 (Kibbe, 2007) and the results are listed in Appendix Table D.1. As the phosphate groups of two nucleotides are bound together, an OH-group is cleaved. This results in a lower molecular weight of Nucleotides in the polymer chain.

TABLE D.1: Molecular weight of nucleotides (Nt)

Base	MW ($gmol^{-1}$)	MW of polymerized Nt ($gmol^{-1}$)
Guanine	363.2	345.2
Cytosine	323.2	305.2
Adenine	347.2	329.2
Uracil	324.2	306.2

D.2 Supporting information B:

The activity in transcription and translation results in an additional ATP-demand, which is related to the growth-independent ATP-maintenance shown in Appendix figure D.1. For transcription, the cost is derived from a nucleotide balance that includes the nucleotides set free by degradation and the nucleotides used in the mRNA synthesis process. The cost for translation follows the course of the active ribosomes. The translation cost exceeds the cost for transcription at least with a factor of 2.5.

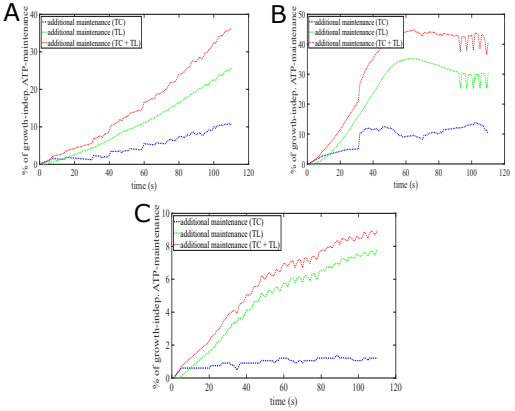


FIGURE D.1: Simulated additional ATP-demand of transcription (TC) and translation (TL) at 25 min (A), 2h (B), 28 h (C)

D.3 Supporting information C:

In figure D.2, D.3 and D.4, the time courses of the clusters for the upregulated genes at 25 min, 2 hours and 28 hours are shown. The red line signals the mean value at the corresponding time point. The values are interpolated linearly. In total, 254 genes were significantly overexpressed with a threshold of a 1.5 fold change at 25 min.

At 25 min process time as it can be seen in Figure D.2 two main pattern occur in the time courses. Cluster two and three show an immediate raise in mRNA levels. In contrast, cluster one and four show a timely delay of 30 s in their expression. Although, it cannot be excluded that the start of synthesis occurs later in reality due to linear interpolation

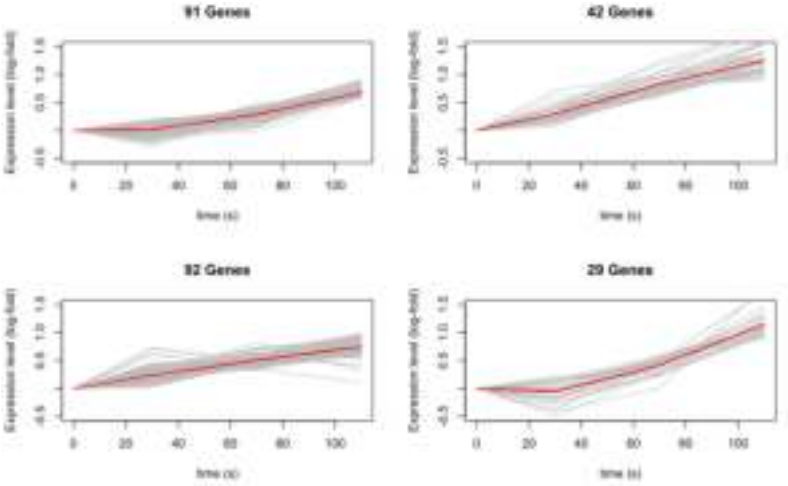


FIGURE D.2: Expression level cluster 1 to 4 at 25 min (left top to bottom right)

between two sampling points. However, assuming that the synthesis starts at 30s, the two clusters show an acceleration in their synthesis rate over the time.

At 2 hours process time in the PFR, in total 276 genes are upregulated. Two genes (*aroF* and *ychH*) are omitted by the algorithm. Exclusion occurs if a gene cannot be added to any cluster without exceeding the maximum intra-cluster distance. In comparison to 25 min time point one cluster shows a decreasing rate of mRNA synthesis during the PFR time. It can be observed, that an additional sigma-factor (σ-24) as well as its antagonist anti-σ-24 *rseA* are activated in the PFR.

At 28 hours, 268 genes are activated of which 6 (*aroF, ecnB, ydhC, ygcG, yedM, yqeC*) were excluded by the algorithm out of the same reason as mentioned above. The three resulting clusters are depicted in Figure D.4. A pattern with delayed increase is still present as can be seen in cluster 1. Furthermore, cluster two, containing almost half of all activated genes shows a steep raise in expression until 50 seconds. After that, the expression level decreases rapidly to 0 or even below its reference value in the STR. The third cluster shows a decreasing rate of synthesis over the PFR, already reaching a steady-state after 50 seconds. In general, it can be noted, that the number of activated genes remains nearly constant while especially the occurring pattern change over time. Moreover only a small group of 80 genes is conserved over the whole process time. To determine the nucleotide

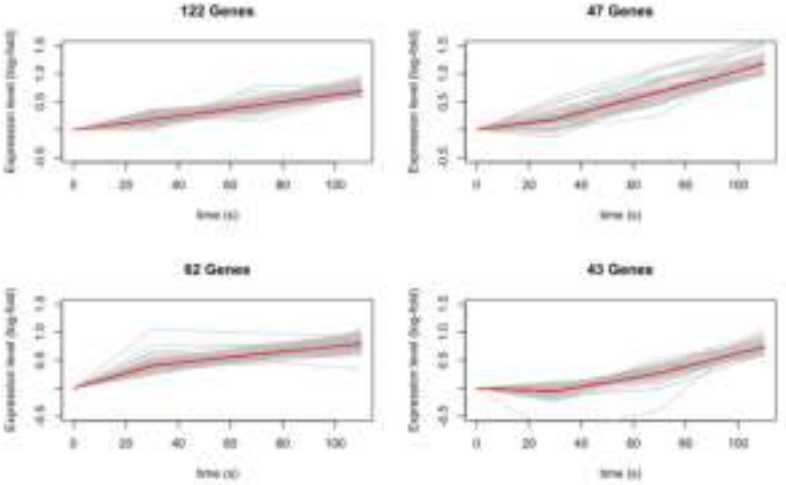


FIGURE D.3: Expression level cluster 1 to 4 at 2 h (left top to bottom right)

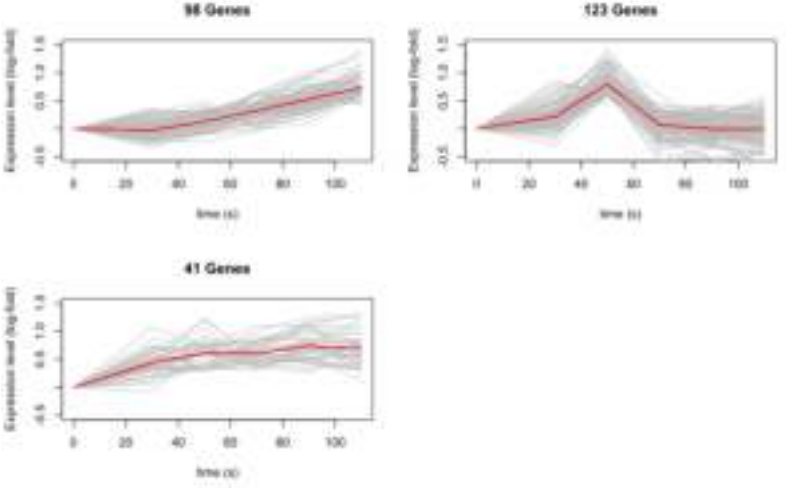


FIGURE D.4: Expression level cluster 1 to 4 at 28 h (left top to bottom right)

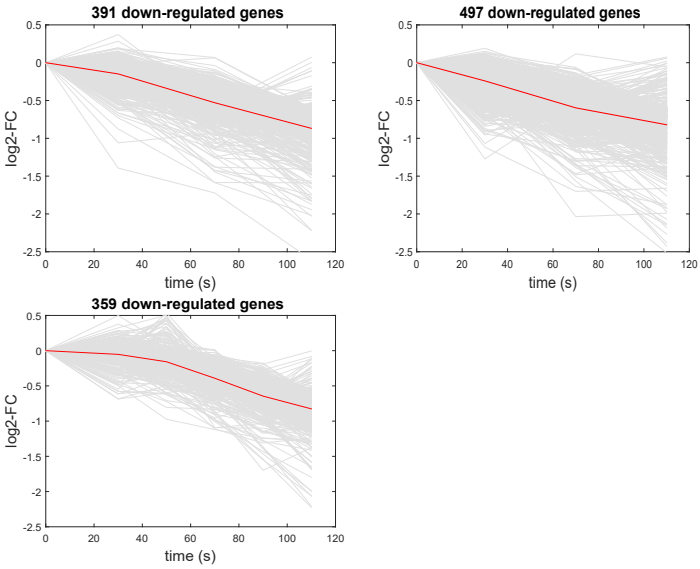


FIGURE D.5: Down-regulated genes. Top left: 25 minutes; Top right: 2h and Bottom left: 28h

balance the downregulated genes are considered (figure D.5). Down-regulated genes for each time point were assigned to one cluster.

D.4 Supporting information D:

A general overview of the properties of the resulting clusters is given in Table D.3, Table D.4, Table D.5. It can be seen, that on average, the reference copy number in the STR is

TABLE D.2: Genes involved in different stress types at three time points

Stimulus	GO-Term	25 min	2 h	28 h
UV-radiation	0071478	2	2	0
Metal ions	0071248	4	4	0
Oxidative stress	0006979	12	16	4
Toxic substance	00097237	3	5	2
Osmotic stress	0006970	0	8	0
DNA-damage	006974	0	25	0
Thermal stress	0009266	0	12	0

< 1. It is concluded that, assuming mRNA occurs only as a complete transcript and not in fractions, that each cell contains a different mRNA pool. Especially Cluster two at 28 h exhibits a very small copy number of only 0.0056 copies per cell. The average copy number increases from 0.22 at 25 min to 0.39 at 2 h. Then, the copy number drops to 0.18 per cell. In contrast, the mean log-fold change ($\log_2$-FC) at 110 s decreases over all three time points from 0.85 to 0.8 and 0.37 at 28 h. It can be observed, that the average gene length of the clusters changes over time, too. While after 25 min, the average gene length of different clusters is close to each other, after 28 h the average length of cluster 2 is almost twice the average length of cluster 1 and more than twice the average length of cluster three.

TABLE D.3: General properties of 25 min clusters

Cluster number	Size	Copy number in STR	Mean gene length	Max. mean $\log_2$-FC at 110 s
1	91	0.18	862	0.68
2	42	0.14	721	1.25
3	92	0.3	937	0.74
4	29	0.21	920	1.16

For comparing the generated clusters of the different time points, a set of indices is calculated. The indices that refer to the compactness and separation of the shown clusters are given in Table D6. The values of the silhouette coefficient, the figure of merit and the intra cluster variance were calculated for each time point separately. Compared with the results of the pure SQBC algorithm, the cluster indices show better values after the use of SQBC-initialized k-means clustering.

TABLE D.4: General properties of 2 h clusters

Cluster number	Size	Copy number in STR	Mean gene length	Max. mean $\log_2$-FC at 110 s
1	122	0.38	980	0.69
2	47	0.16	848	1.18
3	62	0.71	984	0.80
4	43	0.22	657	0.74

TABLE D.5: General properties of 28 h clusters

Cluster number	Size	Copy number in STR	Mean gene length	Max. mean $\log_2$-FC at 110 s
1	98	0.22	717	0.73
2	123	0.0056	1347	0.80
3	41	0.63	635	0.68

TABLE D.6: Cluster indices at different time points

Time point	Silhouette index	Figure of merit	Intra cluster variance
25 min	0.77	0.097	0.017
2 h	0.79	0.078	0.013
28 h	0.61	0.068	0.037

D.5 Supporting information E:

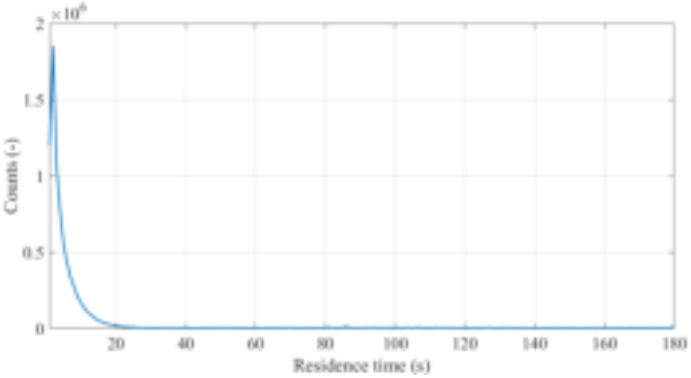


FIGURE D.6: Nonnormalized starvation residence time distribution over 180s simulation time. The weighted mean residence time in the starvation regime is 9.46 s.

D.6 Supporting information F:

As quality criterion for statistical relevance the ergodicity is used. Lagrange trajectories are ergodic if the time average yields the same than the average over the probability space. Therefore, the reactor was spatially discretized into ten cylindrical segments (z_1-z_{10}) and the particles are superimposed at each time step (0-180s). The expected value (black line) is compared to the result and should lie in the confidence level of 95%. As shown in Figure D.7 the superimposed result lies within the confidence level. Additionally, the average glucose concentration of the Lagrangian trajectories is 22.79 mg L^{-1} which is fairly close to the average concentration in the bioreactor of 23.74 mg L^{-1}. Summarizing, only 4% deviation was found which is qualified as a small difference indicating good homogeneous distribution.

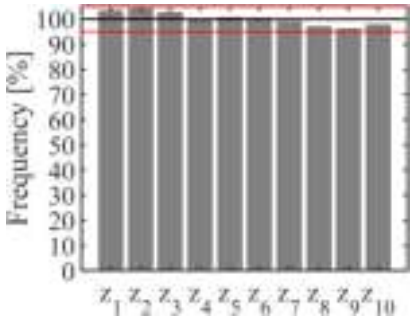


FIGURE D.7: Statistical relevance of Lagrange trajectories. Reactor volume was divided into 10 cylindrical sub-volumes (bottom to top: z_1-z_{10}). Temporal residence of the particles was normalized by corresponding volume segment. For each spatial segment the probability with respect to the expected value (blackline) is shown as the frequency in percentage. Red lines indicate the ± 5 % deviation limit.

D.7 Supporting information G:

Rapid regime transitions occur at the Lagrangian timescale τ_{lg} and reflect subgrid variations of concentrations induced by small eddies. They represent artificial shifts which are unlikely to have metabolic influence and may affect the statistics of large scale fluctuations. Also they can not be address feasibly by conventional scale-down simulators. Therefore an additional filtering step is applied to attain tractable variations in the timescale of turbulent fluctuations: τ_{lg}=0.45$\frac{k}{\epsilon}$. The mean time scale $\overline{\tau_{lg}}$ = 0.13 s, which is in the range of the reaction timescale (0.14 s). To remove such artificial fluctuations a standard moving median filter was used (low pass filter), which smooths high frequency components such as fluctuations caused by the DRW model.

D.8 Supporting information H:

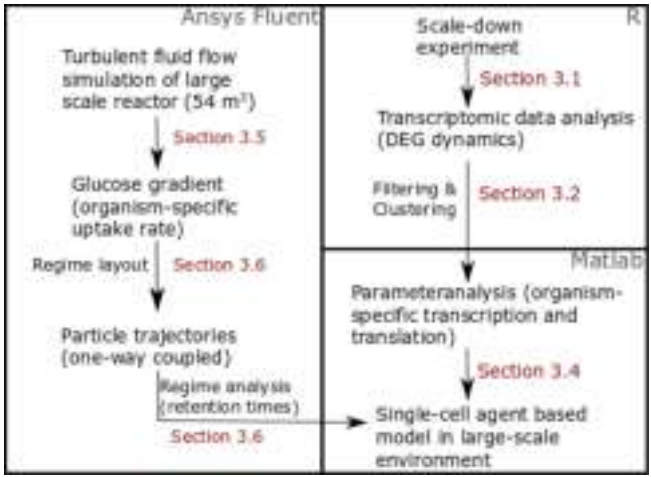


FIGURE D.8: Schematic overview of workflow. The software used is indicated in the upper right corner of each field. In red, reference to manuscript sections is provided offering further details. DEG: Differentially expressed gene.

D.9 Supporting information I:

Repetition of lifelines was used to link particle tracks thereby extending the observation window to 4.5 h process time. Repetiting lifelines bears the intrinsic drawback that cells are 'trapped' on the same trajectories. Alternately, lifelines may be linked randomly. We checked the alternative identifying a high number of artificial shocks imposed on the cells when randomly 'transported' from one end of a lifeline to the beginning of a new trajectory. Randomization of starting points for 5 h process time lead to 48.4 % artificial and unsteady changes in regimes. Applying a filter to limit the new starting conditions within a radius of 1 m around the end-point (after 180 s), still 14.8 % of instantaneous and artificial shifts occurred. Using repeating lifelines regime changes reduced to 10.3 %. Accordingly, we decided for the repetition of the lifelines to minimize the effects of artifacts.

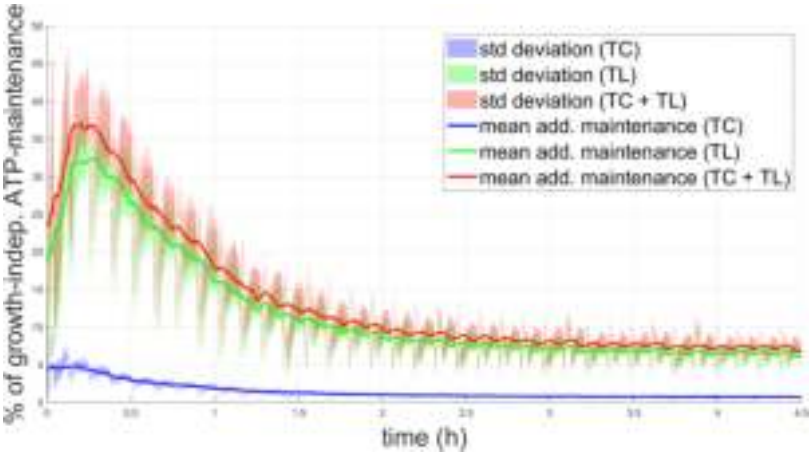


FIGURE D.9: ATP demands of a population comprising 60,000 'newborn', not preconditioned cells in a 54 m^3 reactor monitored over 4.5 hours process time. Courses for mean transcription (blue: TC), translation (green: TL) and the sum of both are depicted (red: TC+TL). Shaded areas display standard deviation.

D.10 Supporting information J:

To evaluate the impact of simulation time additional studies simulating 460 s were performed using the high performance computing cluster at the High Performance Computing Center Stuttgart (HLRS, University of Stuttgart). It was found that cellular ATP demands rise to a maximum of 47 % of growth decouple maintenance. After around 3.7 h cells adapt to environmental changes but still have to bear additional 7 % ATP-demand limiting biomass-specific productivity (figure D.9). The weighted mean residence time is 48.86 s (figure D.10) leading to a broader ATP spectrum and faster adaptation compared to 180 s simulation time.

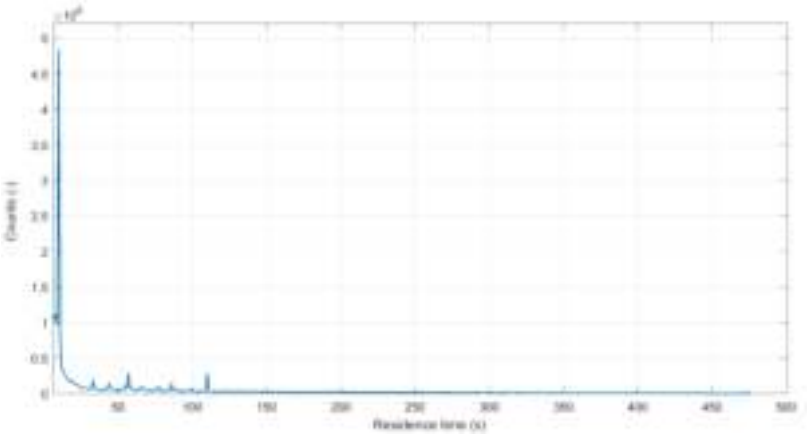


FIGURE D.10: Non-normalized starvation residence time distribution over 460 s simulation time for 60,000 particles. The weighted mean residence time in the starvation regime is 48.86 s.

D.11 Supporting information K:

The characteristic time τ_S for consuming substrate (c_S) may be used as a traditional engineering approach for regime analysis in bioreactors. τ_S is estimated by dividing substrate supply by consumption rates. The latter considers Monod growth kinetics as follows:

$$\tau_S = \frac{c_{S,ave}}{X * Y_{XS} * \mu_{max} * \frac{c_{S,ave}}{K_S + c_{S,ave}}} \tag{D.1}$$

X is biomass concentration and $c_{S,ave}$ is the average substrate concentration in the reactor. Applying eq. D.1 leads to a characteristic substrate consumption time of 63 s, which is smaller than the mixing time ($\tau_{mix} = 4\tau_{circ} = 77\ s$; τ_{circ} : circulation time). Accordingly, substrate gradients are likely to occur. Further analysis may even consider K_S as a critical threshold value deriving reaction volumes with $c_S > K_S$ and $c_S < K_S$ thereof. As such, bioreactors may be divided into several compartments that may serve as basis for scale-down studies. Noteworthy, such approaches assume average values, for instance of P/V, which may differ severely from spatial distributions calculated by CFD studies.

Appendix E

Bubble Size Distribution and Oxygen Mass Transfer in Absence and Presence of Safflower oil

The goal of this project was to determine the influence of safflower oil on bubble surface tension, size distribution and oxygen mass transfer coefficient in a miniaturized industry reactor. Additionally, single phase CFD simulations were extended to multiphase to investigate the distribution of oil in a 3 m^3 reactor.

E.1 Material and Methods

E.1.1 Instruments

A list of all instruments used in this work is provided in Appendix E table E.1.

TABLE E.1: Instruments

Instruments	Manufacturer
Benchtop shaker AK85	Infors AG
Bioreactor KLF 2000 (3.7 L)	Bioengineering
Centrifuge 5430 R; rotor F-35-6-30	Eppendorf
HAMEG®HMP 4040	Rohde&Schwarz
Mass flowmeter Model 3585	Analyt MTC Messtechnik GmbH
Oxygen exchange cap OEC-PSt3-NAU-OIW	PreSens
Oxygen Probe OxyBase WR-RS485-L5-OIW	PreSens
Sony®Alpha 7 camera	Sony
Temperature sensor PT-100	Bioengineering
Tensiometer BP100	Krüss

E.1.2 Chemicals

A list of chemicals used in this work is provided in Appenidx E table E.2.

TABLE E.2: Chemicals

Chemical	Manufacturer
Ammonium sulphate, $(NH_4)_2SO_4$, $\geq$ 99%	Carl Roth GmbH&Co. KG
Potassium dihydrogen phosphate, KH_2PO_4, $\geq$ 98%	Carl Roth GmbH&Co. KG
Potassium hydrogen phosphate, di-, K_2HPO_4, $\geq$ 98%	Carl Roth GmbH&Co. KG
Citric acid, $C_6H_8O_7$	Carl Roth GmbH&Co. KG
Yeast extract	BD
Glucose, α-D(+)-, monohydrate, $\geq$ 99.5%	Carl Roth GmbH&Co. KG
Thiamine	BD
Hydrochloric acid, HCl, $\geq$ 32%	Sigma Aldrich Co. LLC.
Magnesium sulphate heptahydrate, $MgSO_4$, $\geq$ 99%	Carl Roth GmbH&Co. KG
Iron(II) sulfate heptahydrate, $Fe(II)SO_4 \cdot 7H_2O$, $\geq$ 99.5%	Carl Roth GmbH&Co. KG
Manganese(II) chloride tetrahydrate, $MnCl_2 \cdot 4H_2O \geq$ 99%	Carl Roth GmbH&Co. KG
Zinc sulphate septahydrate, $ZnSO_4 \cdot 7H_2O \geq$ 99%	Sigma Aldrich Co. LLC.
Copper(II) chloride dihydrate, $CuCl_2 \cdot 2H_2O$	Merck Chemicals GmbH
Calcium chloride dihydrate, $CaCl_2 \cdot 2H_2O$	Merck Chemicals GmbH
Boric acid, $H_3BO_3 \cdot 2H_2O$	Merck Chemicals GmbH
Sodium molybdate dihydrate, $Na_2MoO_4 \cdot 2H_2O$, $\geq$ 99.5%	Sigma Aldrich Co. LLC.
Sudanschwarz, $C_{29}H_{24}N_6$	Carl Roth GmbH&Co. KG

E.1.3 Software

A list of software used in this work is provided in Appendix E table E.3.

TABLE E.3: Software

Software	Manufacturer
LabVIEW	National Instruments, USA
Matlab R2019b	The MathWorks, Inc.

E.1.4 CGXII Complex Media

Stock solutions of $MgSO_4$, $FeSO_4$ and Thiamine HCl were made by dissolving 200 g $MgSO_4$ in 0.8 L dH_2O, 72 g $FeSO_4$ in 0.05 L dH_2O and 6.5 g Thiamine HCl in 0.005 L dH_2O. Glucose stock solution contained 550g glucose-monohydrate in 1 L dH_2O, which was autoclaved at 121 °C and 2 bar for 20 min. Complex media composition for bioreactor experiments are summarized in table E.4. Complex media for experiments in the

TABLE E.4: Composition of complex medium

Component	Stock concentration [g L^{-1}]	Final concentration [g L^{-1}]
$(NH_4)_2SO_4$)	7.2	216
KH_2PO_4	12	360
K_2HPO_4	4.58	137.4
$C_6H_8O_7$	1	30
Yeast extract	10	300
Glucose	500	600
Thiamine HCl	0.13	4.03
$MgSO_4$	2.49	3.43
Trace elements stock solution (1000x)		120
Trace elements stock solution composition (1000x)		
Fe(II)$SO_4 \cdot 7H_2O$	16.4	1.44
$MnCl_2 \cdot 4H_2O$	10	0.01
$ZnSO_4 \cdot 7H_2O$	0.18	0.002
$CuCl_2 \cdot 2H_2O$	0.18	0.001
$CaCl_2 \cdot 2H_2O$	1.47	0.04
$H_3BO_3 \cdot 2H_2O$	1	0.0005
$Na_2MoO_4 \cdot 2H_2O$	0.08	0.002

glass bioreactor consisted of 30 g L^{-1} glucose, 360 g L^{-1} KH_2PO_4, 137.4 g L^{-1} K_2HPO_4, 216 g L^{-1} $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. The composition of trace element stock solution was 4.175 g L^{-1} Fe(II)$SO_4 \cdot 7H_2O$, 0.002 g L^{-1} $ZnSO_4 \cdot 7H_2O$, 0.01 g L^{-1} $MnCl_2 \cdot 4H_2O$, 0.002 g L^{-1} $ZnSO_4 \cdot 7H_2O$, 0.04 g L^{-1} $CaCl_2 \cdot 2H_2O$ and 3.43 g L^{-1} $MgSO_4 \cdot 7H_2O$. Stock solutions of salts, trace elements, sugars and safflower oil were autoclaved separately and stored at 4 °C. All compounds were combined just before the experiments to prevent potential aging of media. According to Li, Han, et al. (2020), safflower oil remains unaffected by autoclavation.

E.1.5 Oil Coloring

For coloring the organic phase the solely fat-soluble pigment Sudan black (Carl Roth GmbH) was used. One spade tip of pigment was mixed with 30 mL of oil. This step increases contrast to distinguish between bubbles, oil droplets and background in the mixture.

E.1.6 Bioreactor Experiments

Experiments were conducted in a 3.7 L glass stirred tank (STR) (KLF 2000, Bioengineering, Switzerland) with a working volume of 2.45 L and no overpressure applied. The tank was

equipped with four baffles and three impellers (two A340 impellers + one 6-blade Rushton (setup 1, see figure E.1); three A340 (setup 2, see figure E.2)) with identical spacing. Gassing was enabled by a pipe sparger close to the reactor bottom. Spacing of stirrers, other reactor components and liquid filling height were exactly as displayed in figure E.5 and table E.5. Aeration was set to 0.1 or 0.2 vvm and controlled by a mass flowmeter (Model 3585, Analyt MTC Messtechnik GmbH, Germany).

FIGURE E.1: Schematic illustration of three A340 impeller represeting impeller setup 1.

FIGURE E.2: Schematic illustration of two A340 and one rushton impeller represeting impeller setup 2.

Safflower oil was added up to 10 (V/V)% at the beginning of the process and was dispersed for 12 minutes before measurements to reach a stable droplet size distribution. The dissolved oxygen (DO) was measured as percentage of air saturation by an oxygen probe (OxyBase WR-RS485-L5-OIW, PreSens). Measurements were performed as triplicates in oil-in-water and oil-in-medium dispersions. Sampling was enabled by a manual valve. Temperature was kept constant at 37 °C and monitored by a PT-100 sensor (Bioengineering, Switzerland) and controlled through direct contact to an electrical heating element and a water-cooled cooling stick. The LabView® 2009 SP1 software (National Instruments, USA) was used for data recording and process control.

E.1.7 Determination of the Volumetric Oxygen Mass Transfer Coefficient

In this study, the volumetric mass transfer coefficient $k_L a$ was determined with the dynamic gassing out method. The liquid was deoxygenated by replacing air with pure nitrogen. After deoxygenation, air was turned on again until the liquid was saturated. Concomitantly, the oxygen profile was monitored by an oxygen probe (OxyBase WR-RS485-L5-OIW, Presens) with the oxygen exchange cap OEC-PSt3-NAU-OIW. Determination of $k_L a$ was performed via integration of equation E.1 resulting in a linearized representation displayed in equation E.2. In this representation $k_L a$ equals the slope of the line.

$$\frac{dc_{O_{2,L}}}{dt} = k_L a(c^*_{O_{2,L}} - c_{O_{2,L}}) \tag{E.1}$$

$$ln\left(\frac{c^*_{O_{2,L}} - c_{O_{2,L,t_2}}}{c^*_{O_{2,L}} - c_{O_{2,L,t_1}}}\right) = -k_L a(t_2 - t_1) \tag{E.2}$$

with $c_{O_{2,L}}$ as oxygen concentration and $c^*_{O_{2,L}}$ as oxygen saturation concentration. Dynamic oxygen mass transfer experiments were partially conducted by Faiß (2020) under supervision of this thesis author.

Measuring Oxygen Mass Transfer dynamically without Microorganisms

In aerobic fermentations oxygen supply to the organisms must be ensured. As a consequence of very low solubility of oxygen in culture broth, oxygen limitation may occur quickly. Therefore oxygen transfer rate, see equation E.1, is one of the most important parameters in biotechnology. This formula contains the volumetric mass transfer coefficient $k_L a$ which comprises the volumetric interfacial gaseous-liquid surface area a and the mass transfer coefficient k_L. This coefficient only takes the liquid side of the phase boundary into consideration. When the bubble diameter is known, mass-transfer coefficient can be calculated using the penetration theory proposed by Higbie (1935):

$$k_L = \sqrt{\frac{D_{O2}}{\pi}}\left(\frac{\rho\epsilon}{\mu_{lam}}\right)^{0.25} \tag{E.3}$$

Equation E.3 is based on Kolmogorov's Length scale of isotropic turbulence with ϵ as turbulent dissipation rate, μ_{lam} as laminar dynamic viscosity and D_{O_2} as diffusion coefficient. The mass flux depends on the interfacial area a which is given by the following relation:

$$a = 6\frac{\alpha}{d_b} \tag{E.4}$$

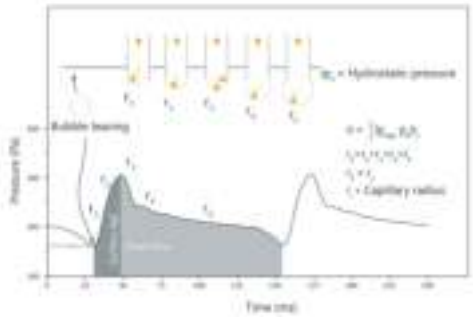


FIGURE E.3: Dynamic surface tension measurement principle. Pressure characteristics depend on curvature radius (r_1 - r_5) and surface age. Capillary radius is indicated as r_c. Figure was kindly provided by KRÜSS, 2020.

where α is the gas hold-up and d_b the bubble diameter, which equals the Sauter diameter d_{32} in CFD simulations.

Since it is difficult to measure each value individually, k_La is measured as a product. It can be interpreted as the reciprocal of the oxygen transfer time (Garcia-Ochoa et al., 2009). The k_La is dependent on geometry parameters (e.g: ratio of tank height to tank diameter or impeller diameter to tank diameter, stirrer type), process parameters (agitation and aeration rate, pressure and sparger type) and liquid properties (density, viscosity, surface tension and coalescence properties) (Van't Riet, 1979).

Probe response time

With a probe response time $t_{63\%}$ = 3.67 s below the critical response time of $t_{63\%,crit}$ = 6.99 s based on k_La_{max} = 103.10 h^{-1}, delay time can be neglected according to the criteria established by Zlokarnik (2013).

E.1.8 Determination of Dynamic Surface Tension

The surface tension of different safflower oil volume fractions (0.2 % and 10 %) in complex glucose medium and water was measured using a dynamic surface tension method. A bubble pressure tensiometer (BP100, Krüss GmbH, Germany) was used and each sample was measured as triplicate. The measurement setup is displayed in figure E.3.

At the beginning of the measurement, a capillary was immersed 2 mm in the dispersion, thereby a magnetic stirrer prevented accumulation of the surfactant at the liquid

surface or capillary. The radius of the capillary was determined by a reference measurement with water preceding every measurement series. Bubbles were formed by airflow through the capillary with radius r_c. The internal bubble pressure depends on the curvature radius (r) and surface tension σ according to the Young-Laplace equation ($p = \frac{2\sigma}{r}$). When a gas bubble is produced at the tip of the capillary, the radius initially decreases until r equals r_c and a pressure maximum p_{max} is reached (see figure E.3). At this point the surface tension can be calculated according to:

$$\sigma = \frac{(p_{max} - p_0)r}{2} \tag{E.5}$$

with p_0 as hydrostatic pressure. The measured values corresponds to the surface tension at a certain surface age, the time from the start of the bubble formation to the occurrence of the pressure maximum as indicated in figure E.3. The dependence of surface tension on surface age can be measured by varying the speed at which bubbles are produced.

E.1.9 Power Input

Power input was measured via a stationary reed switch in combination with a magnet attached to the agitator shaft. When the magnet passes the reed switch the electric circuit is closed. Power input is thereby a function of torque T and can be calculated with the corresponding rotational speed.

$$P = 2\pi N(T_L - T_E) \tag{E.6}$$

with T_L as measured torque value in liquid phase and N as agitation speed. As energy losses occur especially at the agitator shaft's O-ring due to friction torque values in an empty tank (T_E) must be considered in the actual power consumption. Measurements were carried out for different agitator speeds ranging from 100 rpm to 700 rpm and gassing rates (0.1 and 0.2 vvm). The LabView® software (National Instrument, USA) was used for data storage and monitoring and adjustment of impeller speed. The mean power draw from the recorded signal was used for further analysis.

E.1.10 Determination of Oil Droplet and Bubble Size Distribution

To determine the oil droplet and bubble size distribution a Sony® Alpha 7 Camera and a glass reactor equipped with A340 and rushton impellers were used as depicted in figure E.4 with detailed reactor dimensions displayed in figure E.5 and table E.5. Two different

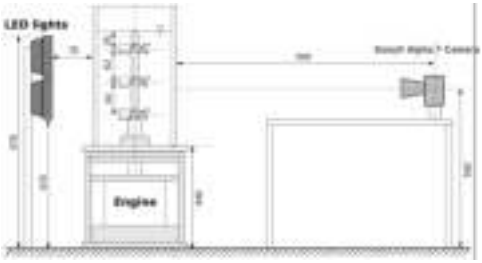


FIGURE E.4: Schematic experimental setup with three A340 impeller. Dimensions are given in mm.

impeller settings (see figure E.1 and E.2) and liquid phases (deionized water; complex glucose medium) and their impact on the oil droplet and bubble size distribution were investigated.

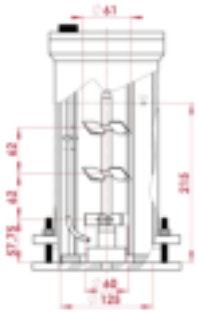


FIGURE E.5: Detailed schematic experimental setup with two A340 and one 6-blade rusthon impeller.

TABLE E.5: Geometric dimensions according to figure E.5

Description	Symbol	Unit
Tank diameter	D	0.125 m
Tank height (liquid level)	H	0.215 m
A340 impeller diameter	d_A	0.061 m
Rushton impeller diameter	d_R	0.060 m
Impeller spacing	ΔC	0.062 m
Off-Bottom impeller clearance	C	0.058 m
Off-Bottom sparger clearance	S_C	0.03 m
Volume	V	0.0245 m^3

Calibration and evaluation of size distributions were conducted as described in Heimann (2017). The evaluation script in Matlab was extended to distinguish between oil droplets and bubbles by integration of an additional size detection factor with which oil droplets are recorded according to their size and color. However, no complete distinction between oil and bubbles can be ensured leading to a potential underestimation of bubble size distribution in the presence of oil since bigger oil droplets may be falsely counted as air bubbles. To reduce the potential overlap in size process parameters should be chosen accordingly. To determine the best process parameters, both phases were first evaluated

separately in corresponding media and impeller setup. Experiments and evaluation of size distributions were performed as part of a bachelor thesis (Fichtel, 2019) supervised by the author of this thesis.

E.2 Results and Discussion

Physical properties of oil-in-water and oil-in-broth dispersions and the impact of safflower oil on bubble size distribution and $k_L a$ were conducted in a glass bioreactor. Numerical simulations were used to investigate oil distribution in an industry size reactor of same geometry as the lab-scale glass reactor.

E.2.1 Power Input

Power input was determined for two different aeration settings (0.1, 0.2 vvm), impeller setups (Setup1: E.1, Setup 2: E.2) and liquids (water, complex glucose media) in the range of 100 - 600 rpm. The results for 0.1 and 0.2 vvm of both impeller setups in water are displayed in figure E.6 on the left hand side.

Same settings were used to monitor power input in complex glucose medium. Results are shown in figure E.6 on the right hand side. Since a maximum volume fraction of 10 % of non-viscous safflower oil was used as a secondary liquid phase, the power input is assumed to remain the same in these mixtures.

The power input increases exponentially with agitation rate for both gassing conditions in setup 2. A diminished increase was visible for setup 1. In comparison to setup 1, setup 2 shows an up to 3.5 times increased power consumption. The additional power

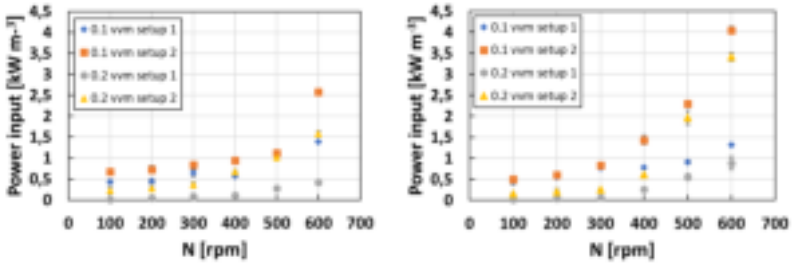


FIGURE E.6: Power input over agitation rate. The gassed power input for different agitation rates (100-600 rpm) in water (left) and complex glucose medium (right) is shown

requirements are due to the increased resistance offered by the fluid to the rotation of the rushton impeller. With an increase in aeration impeller resistance is reduced due to changed fluid densities leading to a lower power consumption (Van't Riet, 1979). Thereby, double the amount of aeration leads to a maximum of 1.6 times reduced power input for setup 2 and 3 times reduced power input in setup 1 in water and 1.2 times for setup 2 and 1.4 for setup 1 in glucose medium (see figure E.6). Although physical properties at 37 °C of broth in terms of viscosity ($0.94 \cdot 10^{-3} Pas$) and density ($1026\,kg\,m^{-3}$) are similar to water at the same temperature (factor: 1.3 of increase in viscosity), power requirements in complex glucose medium are elevated compared to water for high agitation rates. Since the only difference is the liquid used between both experiments, increase in viscosity results in an increase in agitation rate with around the same factor (factor: 1.3).To better compare the obtained values with literature dimensionless power numbers are derived as explained in the next paragraph.

Power consumption values were used to calculate the relevant non-dimensional power numbers (N_P) values, defined as

$$N_P = \frac{P}{\rho_L N^3 D^5} \tag{E.7}$$

where P is agitation power, ρ_L is liquid density, N is agitation speed and D is impeller diameter. Power number as a function of Reynolds number ($\mathrm{Re} = \frac{N\rho D_i^2}{\eta}$) is displayed in figure E.7 for water (left) and broth (right). According to Junker et al. (1998) a N_P of around 5 can be expected for a 6 blade rushton impeller in a baffled reactor in turbulent flow ($\mathrm{Re} > 10^4$) and a N_P of around 0.85 for an A315 impeller, which differ in an additional blade towards the A340 impeller used in this study. For an A340 impeller a slightly lower N_P is expected accordingly. With N_P around 5.1 to 7 for setup 2 and 1.9 to 4.8 for setup 1 this criteria is met in water. For an aeration rate of 0.1 and 0.2 vvm in water using setup 2 a typical rushton profile dominates A340 impellers in turbulent flow ($\mathrm{Re} > 10^4$), whereas the dynamic of setup 1 for equal process parameters resemble the profile of pitched blade impellers (Dickey et al., 1976).

Interestingly, power profiles of impeller setup 2 at 0.1 vvm in medium are elevated compared to water. Since Re number can be interpreted as the ratio between inertia forces and viscous forces, an increase in viscosity leads to a reduction of Re (e.g. viscous forces are dominant in laminar flow). Therefore, a fully turbulent system might not be reached with complex glucose medium in the investigated setup. Laminar flow might still be

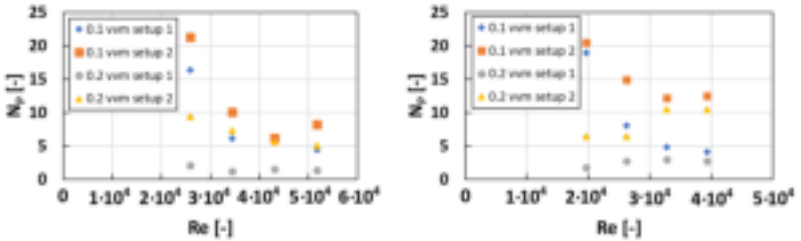


FIGURE E.7: Power number as function of Reynolds number. The gassed power number for different agitation rates (300-600 rpm) in water (left) and complex glucose medium (right) is shown

prominent, where power is directly proportional to dynamic viscosity for a fixed agitator speed (Holland et al., 1995). Additionally, at 0.2 vvm both setup 1 and 2 differ from the anticipated profile as visible in water. With increasing aeration, turbulence is enhanced and a transition from laminar to fully turbulent flow might be reached with reduced impact of viscosity. Further experiments need to be performed to evaluate the investigated profile.

Under turbulent conditions in a baffled tank, the impeller power number is essentially constant and power is proportional to fluid density. Viscosity influences power only in transitional and laminar conditions. In laminar flows where Reynolds number is low, viscosity is dominant. In turbulent flows density is dominant.

E.2.2 Determination of surface tension

The dynamic surface tension of different concentrations of oil in water and broth were measured as introduced in section E.1.8. The results of these measurements are depicted in figure E.8. Measurement series with different volume fractions of oil were conducted in either water (figure E.8 (left)) or complex glucose medium (figure E.8 (right)). Water and glucose medium at 37 °C served as reference in the corresponding measurements. Water with no additives resulted in a constant surface tension of around 70 $\mathrm{N\,m^{-1}}$ irrespective of the surface age, whereas the surface tension of a gas bubble in broth shows a slow decline from 60 to 40 $\mathrm{mN\,m^{-1}}$ in 42 s. As salts are present in glucose medium accumulating at the bubble boundary surface over time, surface tension is continuously reduced until equilibrium is reached (Xu, Nakajima, et al., 2009). Pure safflower oil has a surface tension of 33 $\mathrm{mN\,m^{-1}}$ after 32 s, when thermodynamic equilibrium is reached. Irrespective of the amount of safflower oil added to water (0.2 %, 1 %) the same dynamic of surface tension

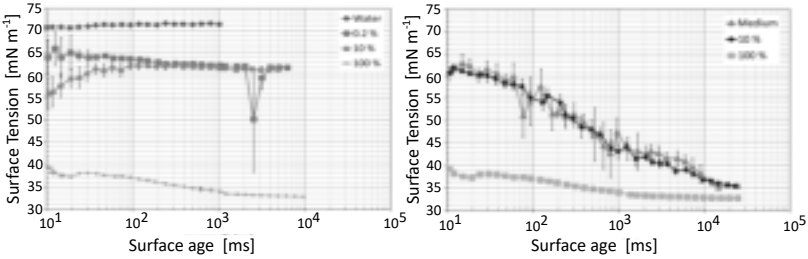


FIGURE E.8: Surface tension of different volume fractions of safflower oil in water (left) and Glucose medium (right)

TABLE E.6: Corresponding sauter diameter to size distribution of bubbles in figure E.9,E.10,E.11,E.12

		100 rpm	200 rpm	300 rpm	400 rpm	500 rpm	600 rpm
Setup 1	Water	0.0042	0.0049	0.0031	0.0022	0.0019	0.0017
	Media	0.0044	0.0048	0.0030	0.0022	0.0019	0.0017
Setup 2	Water	0.0038	0.0029	0.0020	0.0019	0.0016	0.0015
	Media	0.0041	0.0027	0.0019	0.0019	0.0016	0.0015

reduction is visible. In broth however, salts seem to superimpose the effect of safflower oil on surface tension resulting in the same dynamic as pure glucose medium and almost the same surface tension as pure safflower oil.

E.2.3 Separate Size distribution of oil droplets and bubbles

In order to determine size distribution ranges of oil droplets and bubbles and to prepare the evaluation script as described in section E.1.10, both phases were measured separately. Aeration was set to 0.2 vvm during bubble size measurements and no aeration was included when oil droplets were measured. To increase the contrast between oil droplets and its background and to better distinguish between both phases, the oil was stained black as described in section E.1.5. To compare size distributions over different settings the Sauter diameter was calculated as displayed in equation 2.8. For a proper resolution of single oil droplets the volume fraction was reduced from 10 % to 0.2 %. Nevertheless, the effect on surface tension remains the same for both fraction as shown in figure E.8 and thus the effect on oil droplet breakage was considered identical.

Bubble size distribution decreases with increasing rpm. Additionally, distribution becomes narrower for rpm > 300, while for rpm lower than 300rpm bubbles diameters are

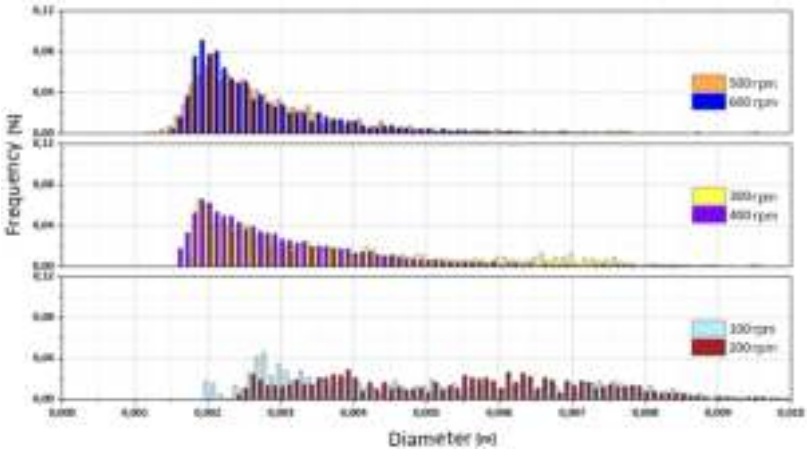


FIGURE E.9: Bubble size distribution in water for different agitation rates (100 - 600 rpm) with 0.2 vvm aeration in setup 1

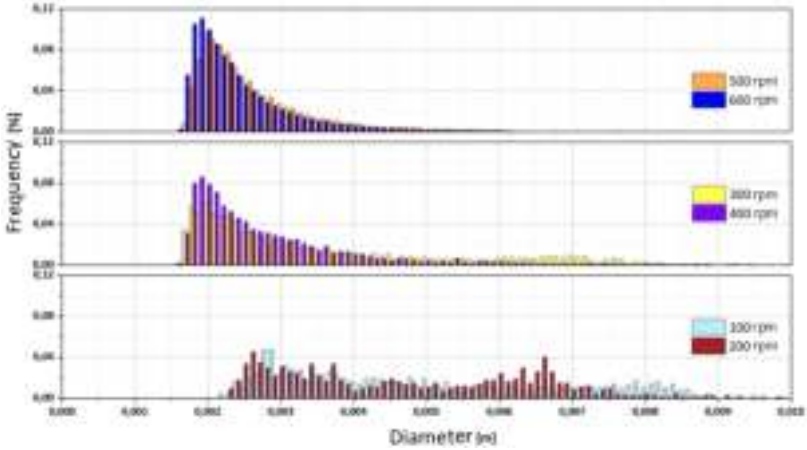


FIGURE E.10: Bubble size distribution in Glucose Medium for different agitation rates (100 - 600 rpm) with 0.2 vvm aeration in setup 1

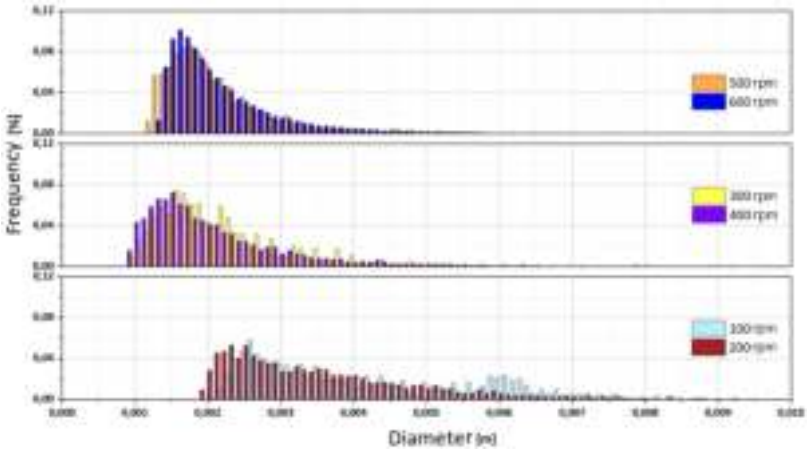


FIGURE E.11: Bubble size distribution in water for different agitation rates (100 - 600 rpm) with 0.2 vvm aeration in setup 2

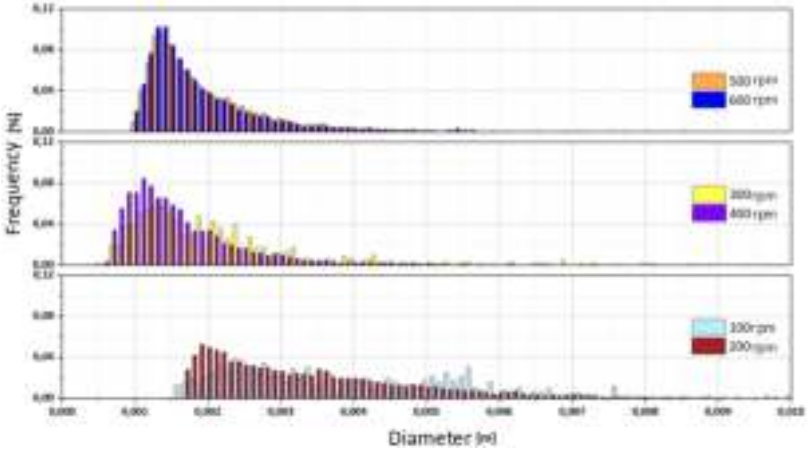


FIGURE E.12: Bubble size distribution in Glucose Medium for different agitation rates (100 - 600 rpm) with 0.2 vvm aeration in setup 2

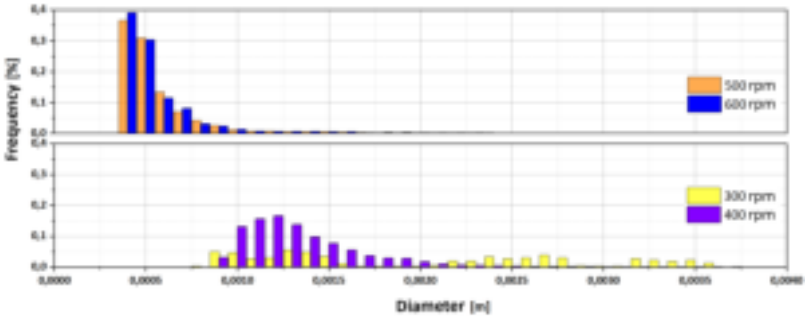


FIGURE E.13: Oil droplet size distribution in water for different agitation rates (200 - 500 rpm) without aeration in setup 1

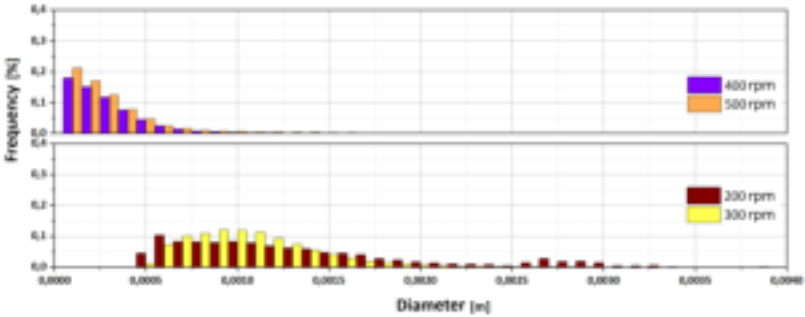


FIGURE E.14: Oil droplet size distribution in Glucose medium for different agitation rates (200 - 500 rpm) without aeration in setup 1

distributed more widely. During all measurements, setup 1 shows smaller diameters compared to setup 2 due to the increased shear stresses at the tip of the rushton impeller resulting in a higher bubble/droplet break-up. Additionally, presence of ions in glucose medium leads to smaller bubbles compared to water (see table E.6) due to zeta-potential reduction of air bubbles leading to reduced surface tension, as shown in section E.2.2.

Evaluation of oil droplet size distribution was not possible for all power inputs measured. For low power inputs oil remained on the surface, whereas for higher turbulences (agitation > 500 rpm) the oil droplets could not be detected due to an extremely fine homogenization. Minimum possible diameter detectable was around 0.0001 m.

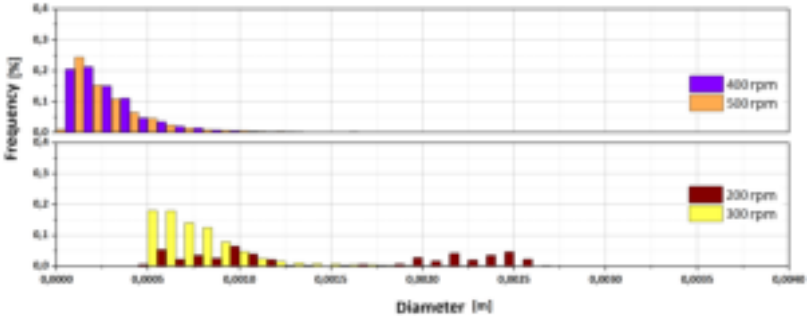


FIGURE E.15: Oil droplet size distribution in water for different agitation rates (200 - 500 rpm) without aeration in setup 2

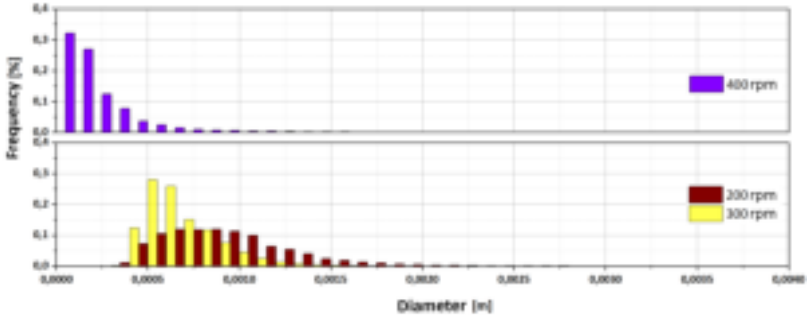


FIGURE E.16: Oil droplet size distribution in Glucose medium for different agitation rates (200 - 400 rpm) without aeration in setup 2

TABLE E.7: Corresponding sauter diameter to size distribution of oil in figure E.13,E.14,E.15,E.16

		200 rpm	300 rpm	400 rpm	500 rpm	600 rpm
Setup 1	Water	-	0.0018	0.0008	0.0005	0.0005
	Media	0.0011	0.001	0.0006	0.0005	-
Setup 2	Water	0.0014	0.0006	0.0006	0.0005	
	Media	0.0008	0.0005	0.0005	-	-

As shown in figure E.14 and E.16 and summed up as Sauter diameter in table E.7, glucose media supports the break-up of oil droplets as already predicted by reduction of surface tension (figure E.8). The Sauter diameter of oil droplets in media is around half compared to water. As stated earlier, surface tension of bubbles decreases during the presence of salts and increases stability by enhancing the structures of the adsorption monolayer and interfacial film (Xu, Nakajima, et al., 2009). Similar mechanisms might be true for oil droplets as their Sauter diameter decreases in glucose media. Oil and bubble size distributions could be best distinguished between 400 and 500 rpm which is why both agitation rates were used for measuring bubble size distribution in the presence of oil.

Smaller droplets/ bubbles enhance species mass transfer by providing increased interfacial area. To examine the effect of oil on volumetric oxygen mass transfer, $k_L a$ values were measured in a lab-scale bioreactor as described in section E.1.6. The results are presented in the following.

E.2.4 Bubble Size distribution with and without oil

After determination of individual size distributions of bubbles and oil droplets, agitation speeds of 400 and 500 rpm with aeration set to 0.2 vvm were used to measure bubble size distribution in the presence of oil. Due to almost identical size distribution in glucose medium, only detailed bubble size distributions in water are shown in the following. For detailed information about bubble size distributions in complex glucose medium see Fichtel, 2019.

TABLE E.8: Oxygen mass transfer coefficient for different oil-in-water volume fractions (0.2 % and 10 %) in setup 1 with 0.2 vvm at 300 rpm

	Reference (w/o oil)	0.2 % oil	10 % oil
$k_L a$ [h^{-1}]	13.1	9.98	5.9

In presence of oil, the number of smaller bubbles increases over all agitation rates in water and broth. Interestingly, albeit bubbles do show a reduced surface tension in oil-in-glucose medium dispersions (35 mN m^{-1}) compared to oil-in-water mixtures (63 mN m^{-1}), the Sauter diameter only differs marginally (see table E.9). Since bubble size distribution is rather broad, slightly reduced bubble sizes are not very well reflected in the Sauter diameter. Although oil does have an impact on bubble size distribution, interactions of oil in broth with other components like yeast extract or salts might be possible. Explanation of bubble size reduction via surface tension might be too simplified in this

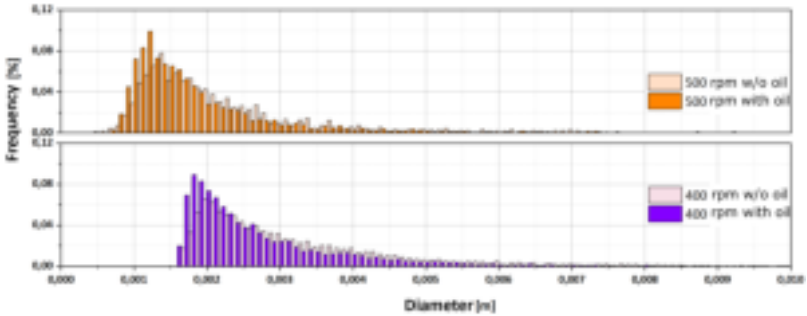


FIGURE E.17: Bubble size distribution in water for different agitation rates (300 - 500 rpm) with 0.2 vvm aeration and 0.2 % oil in setup 1

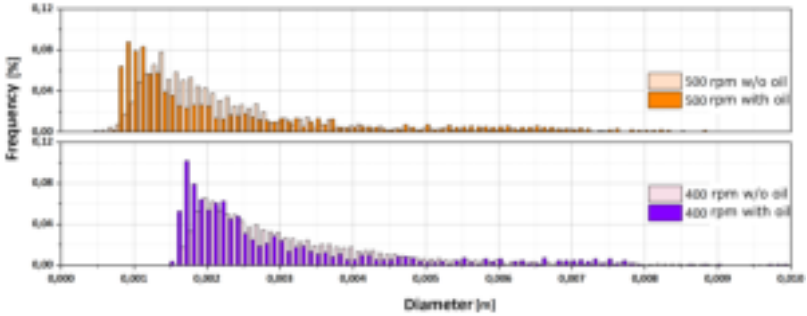


FIGURE E.18: Bubble size distribution in gluocse medium for different agitation rates (300 - 500 rpm) with 0.2 vvm aeration and 0.2 % oil in setup 1

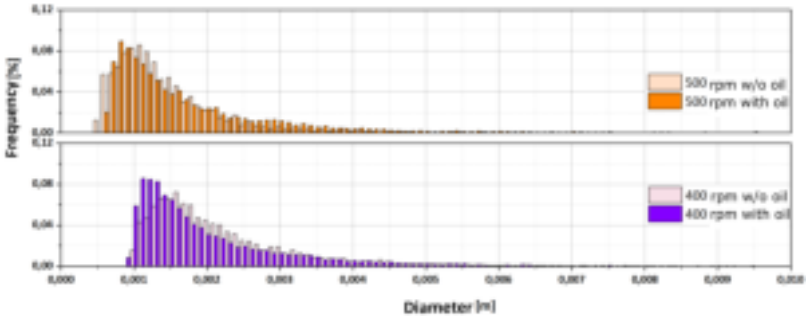


FIGURE E.19: Bubble size distribution in water for different agitation rates (300 - 500 rpm) with 0.2 vvm aeration and 0.2 % oil in setup2

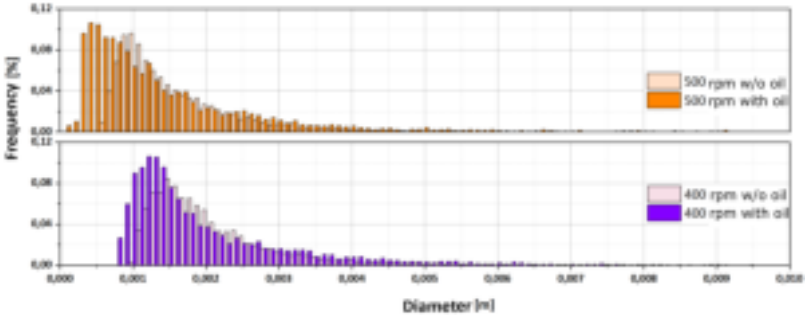


FIGURE E.20: Bubble size distribution in glucose medium for different agitation rates (300 - 500 rpm) with 0.2 vvm aeration and 0.2 % oil in setup 2

TABLE E.9: Corresponding sauter diameter to size distribution in figure E.17,E.18,E.19,E.20 with 0.2 vvm

	Setup 1				Setup 2			
	Water		Glc-Medium		Water		Glc-Medium	
Oil addition	NO	YES	NO	YES	NO	YES	NO	YES
400 rpm	0.0022	0.0021	0.0022	0.0020	0.0019	0.0018	0.0019	0.0018
500 rpm	0.0019	0.0018	0.0019	0.0018	0.0016	0.0015	0.0016	0.0015

case. Interaction of other broth components with oil need to be examined in separate studies.

However, addition of oil shows a positive effect on the Sauter diameter in water and glucose medium for both setups by increasing the interfacial area of air bubbles. To validate the anticipated effect of increase oxygen mass transfer, $k_L a$ measurements were performed and corresponding results are shown in the following.

E.3 Mass transfer coefficients

Since positive tendencies in terms of reduced surface tension and reduced bubble size distribution in the presence of safflower oil could be shown, oxygen mass transfer coefficients were measured next. Two different aeration settings were chosen (0.1 and 0.2 vvm) in combination with various agitation rates (300 - 600 rpm) and 10 (V/V)% safflower oil to determine the influence of oil on $k_L a$. Oxygen probe Modell OEC-PST3-NAU-OIW from Presens was used for all measurements. Probe cap (OEC-PSt3-NAU-OIW) was coated with teflon to prevent accumulation of oil on the sensor surface. Results for impeller setup 1 and setup 2 are displayed in figure E.22 and E.21, respectively. Contrary to the expectations based on the results of previous results (surface tension, bubble size distribution) $k_L a$ values are diminished up to 50 % when oil is added in setup 1 (figure E.22) and up to 20 % in setup 2 for the same gassing rates (figure E.21). No increased values in the presence of safflower oil were detected in neither of the setups and conditions used. Overall, $k_L a$ values are almost twice as high in setup 2 compared to setup 1 due to the improved break-up of bubbles at the rushton impeller located directly at the top of the sparger (figure E.21). Additionally, $k_L a$ in complex glucose medium is slightly higher compared to water in setup 2 due to bubble coalescence inhibition of ions (Henry et al., 2007). Interestingly, glucose medium seem to have a negative effect on $k_L a$ in setup 1, which needs to be investigated further. However, as there is not much opportunity for bubbles to coalesce in this small and highly turbulent system, the effect of salts inhibiting coalescence might be more prominent in larger tanks where different turbulent zones occur.

While the addition of oil decreases surface tension facilitating break-up and resulting in smaller bubbles, the $k_L a$ value does show the opposite effect in all presented studies. Therefore the sole assumption of oil enhancing oxygen mass transfer might be to simple in this scenario. Dumont et al. (2003) and Kaur et al. (2007) reported similar contradictionary results for different studies. The opposing results leads to the assumption that

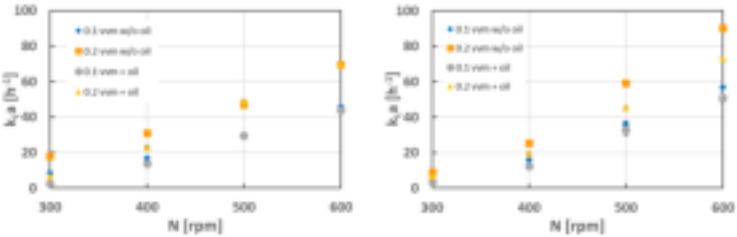


FIGURE E.21: $k_L a$ in Setup 2 measured with different aeration rates with and without oil. Water (left), glucose medium (right)

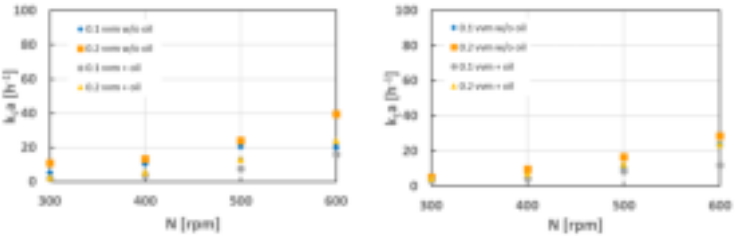


FIGURE E.22: $k_L a$ in Setup 1 measured with different aeration rates with and without oil. Water (left), glucose medium (right)

beneath a positive influence on bubble break-up and reduced surface tension, surface mobility is inhibited by oil leading in turn to a reduction of oxygen mass transfer rate. Hydrodynamics change by suppression of surface mobility as well as interfacial blockage by increasing mass transfer resistance. To further investigation potential of spreading of oil on bubble surface, the spreading coefficient was determined as presented in the next chapter. Studies of Patil et al. (2020) however indicate a positive effect of safflower oil during fermentation processes on $k_L a$. Since producing microorganisms are involved in this study, oil might behave differently in the presence of (by-) products and microorganisms. Additionally, antifoam was added to the fermentation broth over the time course of this process which might also interact with oil. Further analysis is needed to quantitatively describe oxygen mass transfer behavior in the presence of safflower oil. Because of this ambivalent behavior no mechanism could be extracted to mathematically describe the effect of oil on bubble break-up to simulate oxygen mass transfer in industry scale reactors.

Spreading coefficient

With the individual surface tensions (see section E.2.2) measured, the spreading coefficient as explained in chapter 2.9 results in -0.23 mN m^{-1}. Yoshida et al. (1970) showed, that negative spreading coefficients lead to a decrease in $k_L a$ with an increase in oil volume fraction, based on studies with kerosene and paraffin. For safflower oil a similar trend can be observed as displayed in table E.8, where $k_L a$ decreases with increasing volume fraction matching the sense of a negative spreading coefficient. However, according to Brilman et al. (1998) the initial spreading coefficient is irrelevant in well mixed stirred tank fermenter and the actual spreading coefficient appears to be negative for all systems due to oil and bubbles being forced to come together. Since no final conclusion can be drawn from spreading coefficient value, further investigations need to be performed.

E.4 Conclusion

Bubble size distribution and oxygen mass transfer coefficients have been measured experimentally with and without safflower oil to investigate the effect of oil on $k_L a$. Safflower oil reduces surface tension from 71 mN m^{-1} to around 61 mN m^{-1} at thermodynamic equilibrium in water but no additional effect in media was visible. The amount of oil (0.2 or 10 %) volume fraction was thereby irrelevant. Measurements of bubble size distribution are in line with the effect of oil on surface tension. Reduced surface tension in the presence of oil reduces bubble diameters enhancing interfacial area for increased species mass transfer. However, $k_L a$ measurements showed the opposite of the anticipated effect. $k_L a$ values in oil-in-liquid systems regardless of the setup or continuous liquid used never exceeded $k_L a$ values in oil-free systems. This shows that oil has a much more complex effect on mass transfer which can not be solely explained by surface tension reduction. Thorough analysis is needed to quantify these effects. Comparison of impeller setups resulted in improved bubble break-up with simultaneously increasing power input when a A340 impeller is substituted by a rushton impeller above the sparger.

Metabolic Engineering 67 (2021) 75–87

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Engineering of a robust *Escherichia coli* chassis and exploitation for large-scale production processes

Martin Ziegler , Julia Zieringer , Clarissa-Laura Döring , Liv Paul , Christoph Schaal , Ralf Takors *

University of Stuttgart - Institute of Biochemical Engineering, Allmandring 31, 70569, Stuttgart, Germany

ARTICLE INFO

Keywords:
Large-scale
Fermentation
E. coli
Protein production
Deletions
Heterogeneities

ABSTRACT

In large-scale bioprocesses microbes are exposed to heterogeneous substrate availability reducing the overall process performance. A series of deletion strains was constructed from *E. coli* MG1655 aiming for a robust phenotype in heterogeneous fermentations with transient starvation. Deletion targets were hand-picked based on a list of genes derived from previous large-scale simulation runs. Each gene deletion was conducted on the premise of strict neutrality towards growth parameters in glucose minimal medium. The final strain of the series, named *E. coli* RM214, was cultivated continuously in an STR-PFR (stirred tank reactor – plug flow reactor) scale-down reactor. The scale-down reactor system simulated repeated passages through a glucose starvation zone. When exposed to nutrient gradients, *E. coli* RM214 had a significantly lower maintenance coefficient than *E. coli* MG1655 ($\Delta m_s = 0.038\ g_{Glucose}/g_{CDW}/h$, $p < 0.05$). In an exemplary protein production scenario *E. coli* RM214 remained significantly more productive than *E. coli* MG1655 reaching 44% higher eGFP yield after 28 h of STR-PFR cultivation. This study developed *E. coli* RM214 as a robust chassis strain and demonstrated the feasibility of engineering microbial hosts for large-scale applications.

1. Introduction

Large-scale fed-batch bioprocesses often suffer from reduced process performance compared to lab-scale experiments conducted during process development (Bylund et al., 2000; Enfors et al., 2001). The physical and engineering constraints in large-scale reactors inevitably lead to the formation of spatial heterogeneities of relevant process parameters such as nutrient availability, concentrations of dissolved oxygen, carbon dioxide, and pH (Bylund et al., 1998; Cortés et al., 2016). Heterogeneities of nutrient availability are caused by long mixing times of large-scale reactors (Delvigne et al., 2006; Noorman, 2011). Studies employing computational fluid dynamics (CFD) have revealed that in fed-batch processes this typically results in the formation of zones with high nutrient concentrations close to the feeding point and zones depleted of nutrients at the far end of the reactor (Haringa et al., 2017; Kuschel and Takors, 2020). Depending on the mixing time and their position in the reactor cells frequently move through different zones on a timescale of seconds to minutes and cellular regulatory programs ranging from overflow metabolism to starvation responses are repeatedly triggered and shut down (Kuschel et al., 2017). Due to the delay of transcriptional responses, regulatory consequences of stress stimuli may be effective distant from the spot of stress induction which finally creates a heterogeneous population status (Nieß et al., 2017; Zieringer et al., 2020). There is evidence from an increasing number of studies that the performance of many industrial workhorse organisms such as *Escherichia coli*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Saccharomyces cerevisiae* and *Penicillium chrysogenum* is negatively affected when facing process heterogeneities (George et al., 1993; Jonge et al., 2011; Junne et al., 2011; Larsson and Enfors, 1988; Olughu et al., 2020; Vasilakou et al., 2020).

Substantial effort has been made by the scientific community to understand microbial responses to the different zones occurring in large-scale reactors (Lara et al., 2006a, 2006b; Löffler et al., 2016; Olughu et al., 2019). In academic laboratories, the conditions of industrial reactors are commonly simulated using multi-compartment scale-down reactors (Delvigne et al., 2017; Neubauer and Junne, 2010; Takors, 2012). Typically, nutrient pulsing or secondary vessels are employed to deliver a stimulus representative for the conditions under investigation

* Corresponding author. University of Stuttgart - Institute of Biochemical Engineering, Allmandring 31, 70569, Stuttgart, Germany.
E-mail addresses: martin.ziegler@ibvt.uni-stuttgart.de (M. Ziegler), julia.zieringer@ibvt.uni-stuttgart.de (J. Zieringer), clarissa-laura.doering@uniklinik-freiburg.de (C.-L. Döring), liv.paul@gmx.de (L. Paul), christoph.schaal@ibvt.uni-stuttgart.de (C. Schaal), takors@ibvt.uni-stuttgart.de, ralf.takors@ibvt.uni-stuttgart.de (R. Takors).

https://doi.org/10.1016/j.ymben.2021.05.011
Received 8 December 2020; Received in revised form 30 May 2021; Accepted 31 May 2021
Available online 4 June 2021

(Bylund et al., 1999). The design of a scale-down reactor also serves to control the circulation of the microbial population and its residence time in stimulus zones. A commonly used design follows the two-compartment approach comprising a primary stirred tank reactor (STR) coupled to a secondary plug-flow reactor (PFR). While the STR represents the bulk of the fermentation broth, the plug-flow compartment represents a stimulus zone with a defined residence time. Together, the STR-PFR two-compartment reactor enables the study of cellular behavior in heterogeneous environments.

Zones with low nutrient concentration but high oxygen availability occur in reactor segments far away from the feeding point. The effects of such transient starvation conditions on the performance and intracellular regulation of microbial populations can be studied in C-limited scale-down reactors. In the case of *Escherichia coli* K-12 repeated passages of cells through starvation zones were found to negatively impact process performance which could be observed as a reduced biomass yield (Neubauer et al., 1995b). In parallel, regulatory responses such as the stringent response and the general stress response are rapidly initiated (Delvigne et al., 2009; Löffler et al., 2016; Neubauer et al., 1995a; Simen et al., 2017; Sunya et al., 2012). Noteworthy, these cellular responses serve rather long-term than short-term needs and appear to be futile if cells enter zones of nutrient access shortly after the induction of the strategic precaution measure. Transcriptional investigations in a carbon-limited STR-PFR system offered a potential link between futile regulation and reduced process performance: Frequent transcriptional reprogramming was proposed to cause high secondary metabolic costs from aberrant transcription and translation (Löffler et al., 2016). It was estimated that an increased maintenance of up to 30–40% was caused by the transcriptional oscillations and a substantial fraction of this originated from the expression of open reading frames whose products appeared to bestow no apparent benefit in a controlled bioprocess employing standard glucose minimal medium.

The data collected by Löffler et al. (2016) led us to propose a novel design approach for production strains: We reasoned that an intelligently engineered deletion strain might have advantages in conditions that repeatedly induce wasteful expression of process-irrelevant genes. A heterogeneous fermentation with repeated transient starvation could then be a suitable testing environment. The choice of deletion targets would have to be based on the estimated effect of the deletion and be restricted by the requirements of neutrality towards growth and global regulation. The design process differs from previous considerations on the creation of lean-proteome strains in the regard that savings only become apparent due to fluctuating induction (Valgepea et al., 2015). Secondary metabolic costs can traditionally be assessed through Pirt's maintenance coefficient (Pirt, 1965). We hypothesized that the deletion of a suitable set of genes should lead to a reduced maintenance coefficient under scale-down conditions representing starvation zones. The resulting strain could then serve as a base strain for the construction of robust production strains.

We identified deletion candidates matching the defined criteria and constructed a series of deletion strains from *E. coli* MG1655. The final strain of the series, named *E. coli* RM214, was fermented in continuous cultivations in an STR-PFR system simulating starvation zones. *E. coli* RM214 had a significantly lower maintenance coefficient than *E. coli* MG1655 under simulated large-scale conditions. We then characterized *E. coli* RM214 in an exemplary protein production scenario using eGFP as a model product. Compared to *E. coli* MG1655, the deletion strain showed an increased resilience towards the scale-down conditions as evidenced by reduced productivity losses and a higher fraction of producing cells.

2. Materials and methods

2.1. Bacterial strains, media, and buffer solutions

All strains used in this study are listed in Table 1.

Table 1
Bacterial Strains used in this study.

Strain	Genotype/Strain Information	Reference/ Source
Escherichia coli K-12 MG1655	F⁻, λ⁻, *ilvG*⁻, *rfb*-50, *rph*-1 ("wild type" strain, abbrev. WT)	Michalowski et al., (2017)
Escherichia coli DH5α *λpir*	*supE*44, Δ*lac*U169 (Φ80lacZΔM15), *recA*1, *endA*1, *hsdR*17, *thi*-1, *gyrA*96, *relA*1, λpir phage lysogen	Michalowski et al., (2017)
Escherichia coli DH10B pSIM5	F⁻ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lacX*74 *rec*A1 *end*A1 *ara*D139 Δ(*ara-leu*)7697 *gal*U *gal*K λ⁻ *rps*L (Str^R) *nup*G	Datta et al., (2006)
T-SACK	W3110 *araD*<>*tetA-sacB-amp fliC*<>*cat argG*::*Tn*5	Li et al., (2013)
Escherichia coli CD101	MG1655 Δ*flk*	This study
Escherichia coli CD201	MG1655 Δ*flk* Δ*fliA*	This study
Escherichia coli CD202	MG1655 Δ*flk* Δ*fliA* Δ*fliC*	This study
Escherichia coli CD203	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL*	This study
Escherichia coli CD204	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR*	This study
Escherichia coli CD205	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA*	This study
Escherichia coli RM206	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD*	This study
Escherichia coli RM207	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA*	This study
Escherichia coli RM208	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR*	This study
Escherichia coli RM209	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpTCBA*	This study
Escherichia coli RM210	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpTCBA* Δ*yeeL*	This study
Escherichia coli RM214	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpTCBA* Δ*yeeL* Δ*flxA*	This study
Escherichia coli BW3110 pJOE4056.2	W3110 *rhaB*⁻	Wegerer et al., (2008)
Escherichia coli DH5α λpir pJOE4056.2_tetA	*supE*44, Δ*lac*U169 (Φ80lacZΔM15), *recA*1, *endA*1, *hsdR*17, *thi*-1, *gyrA*96, *relA*1, λpir phage lysogen	This study
Escherichia coli K-12 MG1655 rhaB⁻	F⁻, λ⁻, *ilvG*⁻, *rfb*-50, *rph*-1, *rhaB*⁻	This study
Escherichia coli RM214 rhaB⁻	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpT* Δ*yeeL* Δ*flxA rhaB*⁻	This study

(*continued on next page*)

Table 1 (*continued*)

Strain	Genotype/Strain Information	Reference/ Source
Escherichia coli K-12 MG1655 rhaB^{-} pJOE4056.2_tetA	F^{-}, λ^{-}, *ilvG*$^{-}$, *rfb*-50, *rph*-1, *rhaB*$^{-}$	This study
Escherichia coli RM214 rhaB^{-} pJOE4056.2_tetA	MG1655 Δ*flk* Δ*fliA* Δ*fliC* Δ*flgNMABCDEFGHIJKL* Δ*fliEFGHIJKLMNOPQR* Δ*flhEABcheZYBRtaptarcheWAmotBA* Δ*cspD* Δ*aldA* Δ*gatABCDR* Δ*uhpT* Δ*yeeL* Δ*flxA rhaB*$^{-}$	This study

2xYT medium was prepared by autoclaving 16 g/l tryptone, 10 g/l yeast extract, 5 g/l NaCl dissolved in demineralized water. For agar plates 18 g/l agar-agar was added prior to autoclavation. For pH indicator plates 0.03 g/l of neutral red and 10 g/l Rhamnose were supplemented from sterile stock solutions directly before pouring. SOC medium was prepared as described previously (Hanahan, 1983). Agar plates for *tetA-sacB* counterselection were prepared as described previously (Li et al., 2013). If strains with antibiotic resistance markers were cultivated, antibiotics were added to media after autoclavation in the following concentrations: Chloramphenicol 20 μg/ml, Tetracycline hydrochloride 10 μg/ml, disodium Carbenicillin 100 μg/ml.

Minimal media for shaking flask experiments and the precultures for bioreactor experiments consisted of 4 g/l glucose, 3.2 g/l $NaH_2PO_4{\cdot}2H_2O$, 11.7 g/l K_2HPO_4, 8 g/l $(NH_4)_2SO_4$, 0.01 g/l thiamine hydrochloride and 0.2% (V/V) trace elements stock solution. Minimal media for batch cultivation in the bioreactor consisted of 13.4 g/l glucose, 1 g/l $NaH_2PO_4{\cdot}2H_2O$, 2.6 g/l K_2HPO_4, 9 g/l $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. In the experiments with strains carrying pJOE4056.2_tetA for GFP production 10 μg/ml Tetracycline hydrochloride and 1 g/l Rhamnose were supplemented. Towards the end of the batch phase about 100 μl of antifoaming agent Struktol J647 was added to prevent foaming upon glucose depletion. Minimal media for continuous chemostat cultivation in the bioreactor consisted of 13.14 g/l glucose, 1 g/l $NaH_2PO_4{\cdot}2H_2O$, 2.6 g/l K_2HPO_4, 9 g/l $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. In the experiments with strains carrying pJOE4056.2_tetA for GFP production 10 μg/ml Tetracycline hydrochloride and 1 g/l Rhamnose were supplemented. Throughout the chemostat phase 50 μl/h of antifoaming agent Struktol J647 were added continuously to the fermentation medium.

The composition of trace element stock solution was 4.175 g/l $FeCl_3{\cdot}6H_2O$, 0.045 g/l $ZnSO_4{\cdot}7H_2O$, 0.025 g/l $MnSO_4{\cdot}H_2O$, 0.4 g/l $CuSO_4{\cdot}5H_2O$, 0.045 g/l $CoCl_2{\cdot}6H_2O$, 2.2 g/l $CaCl_2{\cdot}2H_2O$, 50 g/l $MgSO_4{\cdot}7H_2O$ and 55 g/l sodium citrate dihydrate. Stock solutions of salts, trace elements and sugars were autoclaved separately, and stock solutions of thiamine hydrochloride and the antibiotics were filter sterilized and stored at 4 °C. All compounds were combined just before the experiments to prevent potential aging of media.

PBS-MgCa for the measurement of eGFP fluorescence and flow cytometry analysis contained 8 g/l NaCl, 0.2 g/l KCl, 1.44 g/l Na_2HPO_4, 0.24 g/l KH_2PO4, 1 mM $MgSO_4$ and 0.1 mM $CaCl_2$. Prior to use PBS-MgCa was filtered with a sterile filter (pore size < 0.2 μm) to reduce particle load (Tomasek et al., 2018).

2.2. Construction of deletion strains

Chromosomal modifications were conducted using recombineering methods that have been comprehensively described and reviewed previously (Murphy, 2016). The *tetA-sacB* cassette and lambda recombineering functions provided by pSIM5 were used to perform chromosomal modifications with base-pair precision (Datta et al., 2006; Li et al., 2013). Deletions of single genes were designed to span the coding sequence only and deletions of operons or larger genomic regions were designed to begin with the coding sequence of the first gene and end with the coding sequence of the final gene. All deletions were verified by sequencing. Table S1 contains an annotated list of primers used in this study and **Supplementary Information S2** a more detailed description of the recombineering method used.

2.3. Construction of GFP production strains

The protein expression system used for the bioreactor fermentations closely resembles previously described systems based on pJOE4056.2 (Wegerer et al., 2008; Wilms et al., 2001). For additional stability, the *bla* resistance cassette from plasmid pJOE4056.2 was exchanged for a *tetA* resistance cassette yielding pJOE4056.2_tetA to enable continuous selective pressure under the conditions of a chemostat. Use of pJOE4056.2_tetA requires induction with the rare sugar rhamnose at low glucose concentrations. Prior to plasmid transformation, we thus inactivated the chromosomal copy of *rhaB* encoding rhamnulokinase in *E. coli* MG1655 and *E. coli* RM214 to yield *rhaB*$^{-}$ strains incapable of utilizing the rare sugar rhamnose. **Supplementary Information S2** contains a more detailed description of the procedure.

2.4. Shaking flask cultivations

For growth experiments glycerol stock cultures strains were streaked on 2xTY agar plates and incubated overnight at 37 °C. For precultures, a single colony was picked to inoculate 15 ml minimal medium in a 50 ml baffled shaking flask and incubated at 37 °C on an orbital shaker set to 130 rpm overnight. On the following morning, an inoculum of the preculture was transferred into 50 ml minimal medium in a 500 ml baffled shaking flask to reach a starting OD of 0.2 and the culture incubated at 37 °C on an orbital shaker set to 130 rpm. Samples were drawn hourly using a fixed needle reaching through the attached cotton plug and a syringe. In all shaking flask experiments the wild type strain *E. coli* MG1655 was cultivated in parallel as a reference and data collected from other strains was normalized to this reference data.

2.5. Bioreactor setup

Bioreactor fermentations were carried out in a two-compartment scale-down reactor. The primary reactor was a stirred tank reactor, and a plug flow reactor was used as the secondary compartment mimicking a starvation zone. The plug flow reactor was connected to the stirred tank reactor only after establishment and sampling of a steady state in the chemostat phase. The technical setup has been characterized previously and includes the modifications described by Ankenbauer et al. (Ankenbauer et al., 2020; Löffler et al., 2016). A schematic overview of the two-compartment reactor is shown in Fig. 1 and **Supplementary Information S2** contains a comprehensive description of the setup.

2.6. Preculture, batch cultivation and continuous cultivation

100 μl of glycerol stock seed culture were directly used to inoculate 300 ml of preculture minimal medium in a 3 l baffled shaking flasks and incubated at 37 °C on an orbital shaker set to 130 rpm overnight. The next morning 160 ml of preculture were used to inoculate the bioreactor complementing the total volume in the bioreactor to 1.6 l fermentation broth. Batch fermentation in the bioreactor ensued at 37 °C. Upon depletion of glucose, indicated by a sharp increase in dissolved oxygen tension, feed and harvest lines were connected. The reactor was refilled with feed medium to 1.6 l broth and a constant feed/harvest rate was established. For GFP production experiments with strains carrying pJOE4056.2_tetA the feed rate was set to 5.33 ml/min corresponding to a dilution rate of 0.2 h^{-1}. For bioreactor cultivations aimed at investigating genomic stability and determining the maintenance coefficient of *E. coli* MG1655 and *E. coli* RM214 the batch phase was shortened, and feed rates were set to 8.00 ml/min, 5.33 ml/min, 2.67 ml/min or 1.33

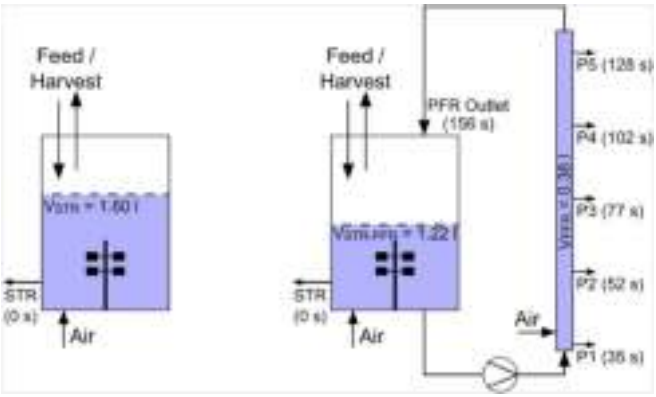


Fig. 1. Reactor Setup. The primary reactor was a standard laboratory reactor operated as a fully aerobic glucose limited chemostat at 37 °C (left scheme). For measurements of the well mixed STR reference state the entire biosuspension was in the primary reactor (V_{STR} = 1.60 l). Scale-down conditions were installed by connecting a secondary plug-flow reactor (PFR). An active pump then constantly circulated V_{PFR} = 0.38 l fermentation broth between STR and PFR reducing the volume fraction in the STR to $V_{STR\text{-}PFR}$ = 1.22 l (right scheme). Labels STR and P1 to P5 designate sampling ports with the respective average residence time of biosuspension after leaving the STR. Fermentations were carried out in two phases each lasting for at least five volumetric residence times: First, a homogenous STR reference state was established, followed by a subsequent heterogeneous STR-PFR phase.

ml/min corresponding to dilution rates of 0.3 h^{-1}, 0.2 h^{-1}, 0.1 h^{-1} or 0.05 h^{-1}. After cultivation for at least five volumetric residence times a reference sample was taken. Then, the plug-flow reactor was connected to the primary reactor via a diaphragm metering pump effectively circulating about one-quarter (380 ml) of the total fermentation broth from the primary reactor through the plug-flow reactor and back into the stirred tank reactor. In the following five to six volumetric residence times samples were taken at predefined time points from the STR and the five PFR ports. Afterwards the fermentation was aborted, and the actual final broth volume measured. This value was used for all volumetric calculations during data analysis.

2.7. Determination of optical density and biomass dry weight

Optical density of fermentation broth appropriately diluted with 0.9% NaCl from the primary reactor was measured in triplicates at 600 nm on a spectrophotometer (Amersham Biosciences/GE Healthcare, Amersham, United Kingdom). For measurement of biomass dry weight quadruplicates of 5 ml broth were centrifuged in weighted glass tubes at 2500 g and 4 °C for 7.5 min. Supernatant was immediately decanted and the pellet washed by resuspending in 5 ml of freshly prepared 150 mM NH_4HCO_3 held at 4 °C. The suspension was centrifuged again, and the washing repeated once. After a final centrifugation, the remaining liquid was decanted carefully, the pellet dried at 105 °C and glass tubes containing dried pellets were weighted again.

2.8. Determination of acetic acid, ammonium and glucose concentrations in fermentation supernatant and feed

5 ml of biosuspension was directly sampled into a syringe connected to a single-use 0.45 µm sterile filter and immediately filtered. The clear supernatant was flash frozen in liquid nitrogen and stored at −70 °C until analysis. Glucose concentration was determined by D-Glucose UV-Test Kit (R-Biopharm, Darmstadt, Germany) and acetic acid concentration by Acetic acid UV-Test Kit (R-Biopharm, Darmstadt, Germany). Ammonium concentration was determined by Ammonium cuvette test LCK 303 or LCK 304 (Hach Lange, Düsseldorf, Germany). At the end of the cultivation feed samples were taken directly from the feed line, flash frozen in liquid nitrogen and processed as described.

2.9. Analysis of total carbon, inorganic carbon and biomass composition

For total carbon and inorganic carbon analysis 0.5 ml biosuspension sample were mixed with 50 µl of 5 M KOH to prevent loss of dissolved carbonate. Then, the suspension was diluted 1:20 with demineralized water, flash frozen in liquid nitrogen, and stored at −70 °C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) to yield the total concentration of carbon and inorganic carbon in the fermenter effluent stream.

To determine biomass composition 1.0 ml of biosuspension was centrifuged at 4 °C and 14000 rpm (20817 g) for 3 min. The supernatant was discarded, the pellet resuspended in 1.0 ml of 0.9% NaCl solution and centrifuged again. The pellet was resuspendend in 5 ml 0.9% NaCl, flash frozen in liquid nitrogen and stored at −70 °C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) and the carbon and nitrogen content of the biomass calculated from these values.

2.10. Measurement of nucleotides

2 ml of biosuspension was sampled directly into 0.5 ml of precooled (<−20 °C) quenching solution and incubated at 6 °C on a shaker for 15 min. Quenching solution consisted of 80 µM EDTA dissolved in 35% (V/V) perchloric acid. 500 µl 1 M K_2HPO_4 was added, and the sample was briefly vortexed. 550 µl 5 M KOH was added and the sample was vortexed again. To remove precipitating potassium perchlorate samples were then centrifuged at 4 °C and 7830 rpm (7197 g) for 5 min. 1.5 ml of supernatant was carefully transferred to new tubes, flash frozen in liquid nitrogen, and stored at −70 °C. Prior to analysis samples were thawed and centrifuged for 10 min at 4 °C and 7197 g. 1 ml of supernatant was transferred to new tubes and their pH adjusted to 6.95–7.05 with 5 M KOH or 35% (V/V) perchloric acid. Samples were centrifuged again for 30 min at 4 °C at 18000 g to remove potassium perchlorate precipitate from neutralization. 500 µl of supernatant were then transferred into RotiSpin Mini 3 kDa MWCO tubes and centrifuged again for 30 min at 4 °C at 18000 g. HPLC analysis was carried out as described previously (Löffler et al., 2016).

2.11. Measurement of eGFP fluorescence

Freshly sampled biosuspension was flash-frozen in liquid nitrogen and stored at −70 °C until analysis. On the day of analysis all samples were thawed and diluted 1:100 with ice-cold PBS-MgCa. 200 µl of diluted sample were transferred into a black 96 well-plate with transparent bottom and lid and the fluorescence (excitation 485 nm, emission 535 nm) was quantified in a SLT SpectraFluor plate-reader (Tecan,

Switzerland). Then, the measured fluorescence values were converted into absolute eGFP concentrations using a calibration curve recorded with purified protein (see **Supplementary Information S2**).

2.12. Flow cytometry analysis

Freshly sampled biosuspension was diluted with PBS-MgCa to yield an OD of approximately 0.04. Diluted biosuspension was passed through a 30 μM CellTrics® filter to reduce particle content and analyzed in a BD Accuri™ C6 Plus Flow Cytometer. The excitation laser had a wavelength of 488 nm and a 533/30 nm emission filter was used to capture GFP fluorescence. Particle signals with a forward scatter height (FSC–H) signal less than 2500 were ignored and 250000 events collected. Events with an eGFP area signal less than 10 were excluded from the analysis to remove dust and cell debris, usually resulting in 235000–249000 remaining events. Cells from events with an eGFP area less than 2000 were defined to form the non-producing population, while cells from events with an eGFP area equal or greater than 2000 were defined to form the producing population. Histograms of all samples can be found in **Supplementary Fig. S3**.

2.13. Genomic DNA sequencing

1 ml of biosuspension was sampled, flash-frozen in liquid nitrogen and stored at −70 °C. On the day of extraction samples were thawn and total DNA extracted with DNeasy Blood and Tissue Kit (Qiagen). Isolated DNA was shipped to and sequenced by the commercial sequencing partner Eurofins Genomics resulting in approximately 5–6 million paired end reads (150 bp) per sample. Data was delivered as fastqc files and assembly of the reads conducted with Unicycler 0.4.8 with the following settings: min contig length 300 bp, min contig coverage 5 (Wick et al., 2017). The obtained contigs were processed with Mauve version 20150226 build 10 using the reference sequence NC_0000913.3 from the NCBI database (Darling et al., 2004). Finally, small nucleotide polymorphisms were detected using snippy (https://github.com/tseemann/snippy) and the output manually examined using Geneious Prime 2020.2.3 (https://www.geneious.com). Supplementary Data S9 contains lists of all SNPs found.

2.14. RT-qPCR

1.5 ml of freshly drawn biosuspension were immediately flash frozen in liquid nitrogen and stored at − 70 °C. Frozen liquid cell suspensions were thawn on ice and 200 μl each were transferred into bead bashing tubes prefilled with 700 μl Lysis buffer. Cells were disrupted with a *Precellys® homogenisator* for 2 × 20 s. RNA was extracted using the *Quick-RNA Fungal/Bacterial Kit* (Zymo Research) following the manufacturer's instructions. The RNA concentrations were measured by Nanodrop. 10 μg RNA each was treated with 2 units *TURBO DNase* (Thermo Fisher Scientific) in 50 μl reactions for 60 min, with additional 2 units enzyme after 20 and 40 min, respectively. RNA from the DNase reactions were purified with *Zymo Clean & Concentrator™-5* (Zymo Research) according to the manufacturer's protocol and were then measured by Nanodrop. cDNA synthesis with *SuperScript® IV reverse transcriptase* (Invitrogen) was carried out according to the protocol for random hexamers as primers. 1 μg RNA was used as starting input for 20 μl reactions, but no RNase inhibitor was added. A no reverse transcriptase control was included. For the qPCR reactions, the cDNA reaction mixes were diluted with 100 μl nuclease free water. 2 μl from all cDNA reactions were pooled together and a dilution series was prepared (1, 1:10, 1:100, 1:1000) for determination of PCR efficiency for each primer pair during each PCR run. For 15 μl reactions 7.5 μl *ORA™ qPCR Green ROX L Mix* (highQu), 0.4 μl forward primer, 0.4 μl reverse primer (f.c. 266 nM, each), 4.7 μl H_2O and 2 μl of diluted cDNA reactions were mixed. *eGFP* was amplified using primers eGFP2-forward and eGFP2-reverse (amplicon length: 248 bp), for *cysG* primers cysG_housekeeping_fwd and cysG_housekeeping_reverse (amplicon length: 197 bp) were used. All reactions were performed as triplicates. Reactions were carried out on a *Biorad CFX96* in 96 well plates. Program parameters were 95 °C, 3 min; 39x (95 °C, 5 s; 59 °C, 15 s; 72 °C, 15 s); 65 °C–95 °C (0,5 °C increment). Data was analyzed with Biorad CFX Manager 3.1. Relative expression of eGFP to cysG was calculated from the c_q numbers measured by the instrument adjusted for amplification efficiency. Relative expressions from time points STR-PFR 25 h and STR-PFR 28 h were normalized to the corresponding STR sample.

3. Results

3.1. Engineering of E. coli deletion strains

Our primary goal was to engineer a series of deletion strains based on *E. coli* MG1655 with physiological advantages under heterogeneous conditions with nutrient depleted zones. Strains would ultimately be assayed in a scale-down reactor consisting of a primary stirred tank reactor (STR) and a secondary plug-flow reactor (PFR) mimicking a starvation zone (Fig. 1).

We began with defining criteria for the choice of handpicked deletion targets: First, only genes that cause relevant metabolic burden in the context of a large-scale bioprocess should be chosen. We thus based our choice of targets primarily on the list of genes with high add-on maintenance under repeated transient starvation published by Löffler et al. (2016) and selected genes with an estimated add-on maintenance > 0.05%. Except for *fliC* none of the chosen genes had an estimated maintenance add-on > 1%, so we expected very little contribution of most single deletions. It was thus clear that multiple deletions would be necessary to achieve reasonably measurable effects. To maximize potential savings, we removed the entire operon if a candidate gene was part of a functionally connected operon. Second, any deletion must not be detrimental to basic growth parameters in glucose minimal medium. In the past, *E. coli* deletion strain series such as the MDS or the MGF series, had suffered from biological fitness losses (Karcagi et al., 2016; Kurokawa et al., 2016). Learning from these studies, any genes involved in primary carbon metabolism or basic cellular functions were outright excluded and we aimed for a highly selective approach with a strictly limited scope. Third, global regulatory programs must be left intact to avoid potential side effects. This included the general stress response, SOS responses and the stringent response. The stringent response had previously been identified as the major repeatedly induced regulatory program but strains with modulated ppGpp availability already exist and have dampened regulatory patterns in nutrient-limited conditions (Michalowski et al., 2017; Ziegler et al., 2020). In this study, one of our goals was to work orthogonally to cellular regulation.

With these criteria in mind, we developed a set of planned deletions containing most parts of the flagellar apparatus, the chemotaxis systems, and multiple other handpicked genes (with addon to maintenance > 0.05%): *cspD*, *aldA*, *flxA*. CspD is a toxin of dispensable function, AldA is irrelevant in glucose-limited medium as its essential function is complemented by PrpC and FlxA is a protein from the Qin prophage. All of these genes are non-essential (Baba et al., 2006). Using lambda recombineering with the *tetA-sacB* cassette we sequentially engineered the strains starting from *E. coli* MG1655 until completion of the final strain *E. coli* RM214 (Table 1). We assayed any new deletion strain from the series for its basic growth parameters in shaking flask fermentations cultivating *E. coli* MG1655 as a benchmark in parallel. None of the deletion strains had major advantages or deficits in maximum specific growth rate or biomass yield in glucose minimal medium affirming our choice of deletion targets (Fig. 2). Conducting the genomic deletions required a high number of total passages until *E. coli* RM214 was completed. We sequenced both the genome of *E. coli* MG1655 and *E. coli* RM214 and identified no problematic mutations (**Supplementary Information S2**, **Supplementary Data S9**). As we expected little impact of single deletions, we decided to focus our characterization only on the

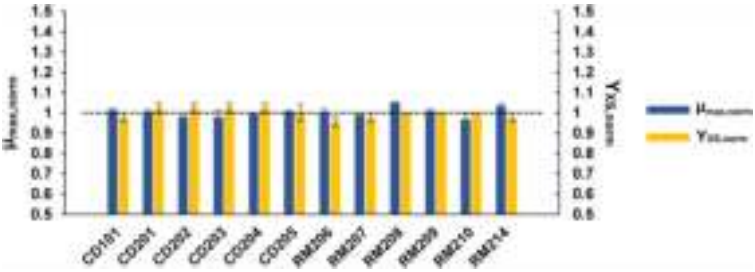


Fig. 2. Basic growth parameters of deletion strains. Deletion strains CD101 to RM214 were cultivated in minimal glucose medium in shaking flask fermentations. The maximum specific growth rate (blue) and biomass yield (yellow) were determined. The parent strain *E. coli* MG1655 was cultivated in parallel, and all data collected normalized to its growth parameters. Error bars indicate SEM (n = 3). The dashed line is a visual aid indicating reference values of 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

final strain of the series, *E. coli* RM214, and compared it to its parent wild-type strain *E. coli* MG1655.

3.2. Maintenance coefficient and genomic stability in scale-down fermentations

To test the initial hypothesis of a reduced maintenance coefficient in heterogeneous conditions and unravel potential benefits of *E. coli* RM214, we cultivated *E. coli* MG1655 and *E. coli* RM214 in two-compartment scale-down fermentations. Continuous chemostat cultivations with two phases were used to enable accurate assessment of fermentation parameters. In the first phase, strains were cultivated in standard well-mixed conditions employing only a STR (Fig. 1, left scheme). After five volumetric residence times this reference state was sampled and the secondary PFR compartment connected to the STR. A diaphragm metering pump then continuously circulated about one-fourth of the fermentation broth from the STR through the PFR and back into the STR. As feeding occurred only in the STR, the PFR simulated repeated passages of fractions of the population through a starvation zone (Fig. 1, right scheme). After continued cultivation for another five volumetric residence times the new STR-PFR steady state was sampled. Therefore, the total process time always exceeded ten volumetric residence times.

We cultivated *E. coli* MG1655 and *E. coli* RM214 at four different dilution rates (0.05 h^{-1}, 0.1 h^{-1}, 0.2 h^{-1}, 0.3 h^{-1}) each. We measured biomass concentrations in the well mixed STR reference state and during the heterogeneous STR-PFR phase. *E. coli* RM214 had a slightly increased biomass yield on substrate, especially under STR-PFR conditions and at D = 0.05 h^{-1}. We estimated Pirt's maintenance coefficient m_s of both strains by linear regression of Y_{XS}^{-1} vs D^{-1} (Fig. 3). We found no statistically significant differences under well mixed STR conditions, but the maintenance coefficient of *E. coli* RM214 was significantly lower than that of *E. coli* MG1655 under STR-PFR conditions (Δm_s = - 0.038 $g_{Glucose} * g_{CDW}^{-1} * h^{-1}$, $p < 0.05$). Differences in the true biomass yield Y_{XS}^{true} were not significant under any conditions ($p > 0.05$). The results confirm the initial hypothesis and the effectiveness of our tailored deletion strategy for the targeted environment.

We sequenced the strains' genomes from the STR-PFR samples from all fermentations to investigate potential genomic instability that may have influenced the observations due to the long fermentation time (>200 h at D = 0.05 h^{-1}). In *E. coli* MG1655, we found SNPs in *insH5* in samples from all dilution rates but no other mutations. We also found SNPs in *insH5* in all samples from *E. coli* RM214 and additional mutations in *ycfK* and *stfE* of the inactive e14 prophage (**Supplementary Data S9**). Apart from these minor alterations, the strains were remarkably stable. They showed no accumulation of mutations in any regulatory genes or genes involved in central metabolism confirming that the engineered deletions bestowed the reduced maintenance coefficient to *E. coli* RM214.

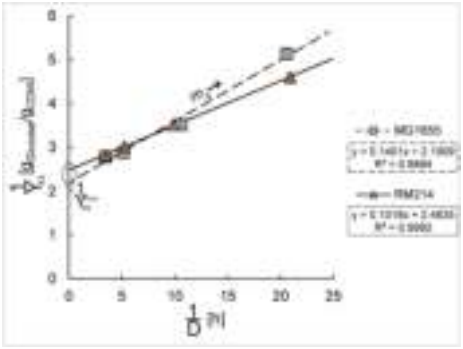


Fig. 3. Determination of maintenance coefficients under heterogeneous STR-PFR conditions. *E. coli* MG1655 (grey squares, dashed line) and *E. coli* RM214 (orange triangles, solid line) were cultivated in the STR-PFR system (glucose limited chemostats, D = 0.05 h^{-1}, 0.1 h^{-1}, 0.2 h^{-1}, 0.3 h^{-1}). Maintenance coefficients m_s (slope) and true biomass yields Y_{XS}^{true} (intersection) were determined from the linear regression of data points. The difference of the maintenance coefficients is statistically significant (Δm_s = - 0.038 $g_{Glucose} * g_{CDW}^{-1} * h^{-1}$, $p < 0.05$). Error bars indicate technical standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.3. Construction of eGFP production strains

Based on these encouraging findings we hypothesized that *E. coli* RM214 should better withstand the stressful conditions of an exemplary heterogeneous production scenario including transient starvation than its ancestor strain *E. coli* MG1655. We chose to produce eGFP as an easily measurable proxy for industrially relevant intracellularly accumulated proteins such as insulin varieties or other biopharmaceuticals commonly produced in *E. coli* (Baeshen et al., 2014, 2015).

A suitable expression system to produce proteins in glucose-limited fermentations is the rhamnose-inducible expression system from pJOE4056.2 (Wegerer et al., 2008). Expression from the rhamnose promoter occurs in the presence of non-toxic rhamnose and is enhanced

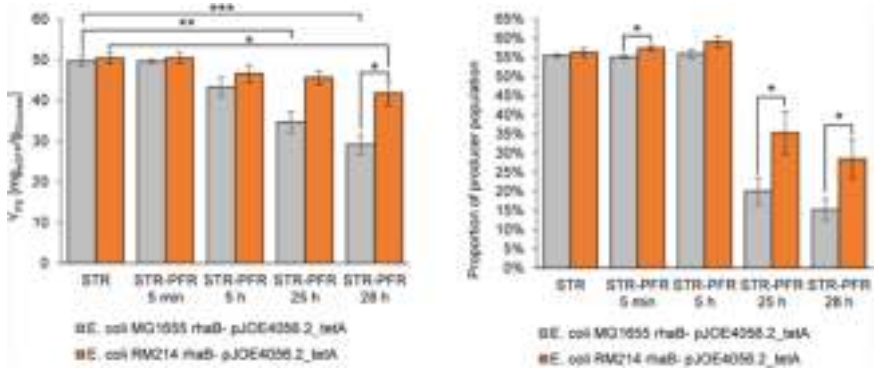


Fig. 4. EGFP yield on substrate and proportion of cells with high eGFP content. *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (grey) and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA (orange) were cultivated in the STR-PFR system (glucose limited chemostat, D = 0.2 h^{-1}). Samples were collected from the primary vessel. Error bars indicate SEM (n = 4), statistical indicators: *p < 0.05, **p < 0.01, ***p < 0.001. **Left**: eGFP yield on substrate declines for both strains after PFR connection. Simultaneously, the difference between the strains gradually increases. Statistics: two-tailed t-tests comparing means of a single strain at later time points to the STR mean of the strain; and comparing the means of both strains at each time point to each other. **Right**: The proportion of cells with high eGFP content declines towards the end of the fermentation and is lower for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA than for *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. Statistics: one-tailed t-tests comparing the presumably lower mean of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA to that of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA at each time point. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

by low levels of glucose sensed by cAMP-CRP signaling. However, the use of rhamnose as a stable inducer requires the absence of rhamnose catabolism (Wilms et al., 2001). We therefore inactivated the chromosomal copy of *rhaB* by replacing the original gene in *E. coli* MG1655 and *E. coli* RM214 with an inactive frameshift copy from *E. coli* BW3110 by recombineering with the *tetA-sacB* cassette. The resulting strains were termed *E. coli* MG1655 rhaB$^-$ and *E. coli* RM214 rhaB$^-$. The absence of rhamnose catabolism was additionally confirmed by streaking the strains on 2xTY pH indicator agar plates containing Rhamnose. *E. coli* MG1655 rhaB$^-$ and *E. coli* RM214 rhaB$^-$ formed white colonies meaning that no acidification of the medium caused by rhamnose degradation occurred.

We then exchanged the *bla* resistance gene from pJOE4056.2 for the *tetA* resistance gene from *E. coli* T-SACK generating pJOE4056.2_tetA (**Supplementary Fig. S2A**). TetA is a tetracycline exporter and thus enables continuous selective pressure in the presence of tetracycline during prolonged cultivations. Transformation of the rhaB$^-$ strains with pJOE4056.2_tetA yielded *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and E. coli RM214 rhaB$^-$ pJOE4056.2_tetA (Table 1).

3.4. Scale-down fermentations with eGFP production

E. coli MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA were then fermented in quadruplicates each in the STR-PFR scale-down reactor in continuous chemostat cultivations at a dilution rate of D = 0.2 h^{-1}. Heterogeneities were introduced by using the two-compartment STR-PFR reactor in the same setting as described above (Fig. 1). Again, this included a well-mixed STR only chemostat phase, and a subsequent STR-PFR chemostat phase to enable direct observation of the short-term and long-term influence of the nutrient-limited zone.

Under well-mixed STR conditions, we observed no substantial differences between the fermentations of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. They reached comparable cell dry weight and eGFP yield on glucose (Fig. 4). In fact, the strains had virtually identical fermentation and production parameters in any parameter measured (Table 2). The primary product eGFP formed a considerable fraction of the total biomass and we detected only trace amounts of acetate byproduct as expected for glucose-limited fermentations. We also determined the proportion of cells with high eGFP content by flow cytometry and found these to be practically identical for both strains in the STR reference steady-state (Fig. 4). As *E. coli* RM214 was specifically engineered to have advantageous traits in heterogenous fermentations including starvation zones these findings were not surprising and instead proved that our genomic deletions did not interfere with the basic fermentation traits of *E. coli* K-12 strains.

Upon connecting the PFR the process performance of both strains started to decline, but this phenomenon occurred remarkably slower and much less pronounced in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA than in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. Five hours after connection of the PFR both strains still had similar fractions of producing cells and reached comparable biomass concentration. However, first differences in cellular eGFP content and product yield already became apparent. Over the remaining process time production parameters increasingly diverged. After 28 h of STR-PFR continuous cultivation we observed a 43% higher product yield for *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA than for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (ΔY_{PS} = 13 $mg_{eGFP}/g_{Glucose}$, two-tailed *t*-test, p < 0.05). Instead, biomass concentration increased in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA indicating a shift from production to biomass formation (Supplementary Fig. S7A). Noteworthy, we found a linear correlation describing the tradeoff between eGFP production and biomass formation using data from both *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA (Supplementary Fig. S7C and Supplementary Fig. S7D). We suspected that the divergence may be caused by a reduced fraction of producing cells for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA compared to *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA and measured the fluorescence of individual cells by flow cytometry. Similar to the eGFP yield, the proportion of actively producing cells shrank rapidly in *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and more slowly in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA. At the final time point the fraction of producing cells was significantly higher for *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA than for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (one-tailed *t*-test, p > 0.05). To check whether differential expression of eGFP might be responsible for the reduction of eGFP yield in the heterogeneous

Table 2
Fermentation parameters[a] of the eGFP production chemostat processes.

	E. coli MG1655 rhaB⁻ pJOE4056.2_tetA		*E. coli* RM214 rhaB⁻ pJOE4056.2_tetA	
	STR	STR-PFR 28 h	STR	STR-PFR 28 h
c_x $[\frac{g_{CDW}}{l}]$	4.02 ± 0.066	4.28 ± 0.024	4.02 ± 0.053	4.13 ± 0.024
c_P $[\frac{mg_{eGFP}}{l}]$	630 ± 20	370 ± 32	650 ± 11	540 ± 41
Y_{XS} $[\frac{g_{CDW}}{g_{Glucose}}]$	0.315 ± 0.0048	0.3360 ± 0.00075	0.314 ± 0.0032	0.322 ± 0.0034
Y_{PS} $[\frac{mg_{eGFP}}{g_{Glucose}}]$	50 ± 1.3	29 ± 2.4	50 ± 1.2	42 ± 3.0
$c_{Acetate,\ STR}$ $[\frac{g_{Acetate}}{l}]$	0.14 ± 0.040	0.017 ± 0.0072	0.010 ± 0.0061	0.07 ± 0.030
$c_{NH_4^+}$ $[\frac{g_{NH_4^+}}{l}]$	1.64 ± 0.025	1.6 ± 0.12	1.57 ± 0.058	1.52 ± 0.033
q_S $[\frac{g_{Glucose}}{g_{CDW} * h}]$	0.64 ± 0.013	0.60 ± 0.012	0.62 ± 0.020	0.60 ± 0.018
q_P $[\frac{mg_{eGFP}}{g_{CDW} * h}]$	32 ± 1.4	18 ± 1.7	31 ± 1.2	25 ± 2.0
Q_{CO_2} $[\frac{mmol_{CO_2}}{h}]$	73.5 ± 0.70	70.6 ± 0.54	72.6 ± 0.49	69.8 ± 0.64
Q_{O_2} $[\frac{mmol_{O_2}}{h}]$	69.9 ± 0.90	67 ± 1.2	69.8 ± 0.90	65.6 ± 0.44
RQ $[\frac{mol_{CO_2}}{mol_{O_2}}]$	1.05 ± 0.019	1.06 ± 0.013	1.04 ± 0.011	1.06 ± 0.010
eGFP content $[\frac{mg_{eGFP}}{g_{CDW}}]$	158 ± 6.5	87 ± 7.3	161 ± 4.6	130 ± 10
Proportion of producing population [%]	55.4 ± 0.43	15 ± 2.8	56 ± 1.2	28 ± 5.1
D $[\frac{1}{h}]$	0.201 ± 0.0026		0.194 ± 0.0028	

[a] Errors indicate SEM (n = 4).

conditions in general or for the differences between the two strains, we conducted RT-qPCR using the housekeeping gene *cysG* as a reference. However, we found no clear indication for differential expression of eGFP towards the end of the fermentation or between the two strains (Supplementary information S10).

After connection of the PFR we observed alterations in the respiratory parameters of both strains. Initially, cells reacted with a short spike of increased respiratory activity which then dropped rapidly in the following hour. The oxygen uptake rate Q_{O2} and the carbon dioxide formation rate Q_{CO2} recovered over the next two volumetric residence times and then slowly drifted towards new steady states but never reached the initial STR only values (Supplementary Fig. S6). We calculated total carbon balances but the deviations in the respiratory rates caused only minor redistributions between the STR reference status and the STR-PFR 28 h sample (Fig. 5, Table S5A and Table S5B). Apart from small gains in the biomass (CDW) fraction and small reductions in the carbon dioxide formation no major differences occurred. Declining

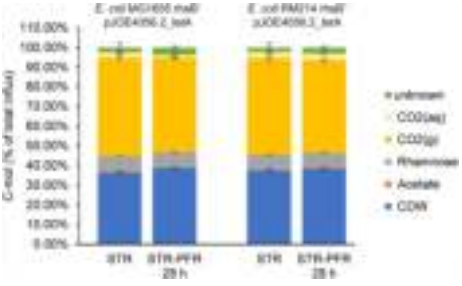


Fig. 5. Carbon Balance. *E. coli* MG1655 rhaB⁻ pJOE4056.2_tetA (grey) and *E. coli* RM214 rhaB⁻ pJOE4056.2_tetA (orange) were cultivated in the STR-PFR system (glucose limited chemostat, D = 0.2 h^{-1}). Columns show efflux fractions of individual substances. Error bars indicate SEM (n = 4). For raw data see Supplementary Tables S5A and S5.B.

productivity was hence accompanied by slightly declining respiration and increased biomass formation. From all collected indications we conclude that the primary factor for loss of productivity of both strains was a restructuring of the biomass composition towards lower eGFP content (Fig. S7B). This is supported by our observations using flow cytometry. The proportion of cells with high eGFP content dropped substantially in the late fermentation stages (Fig. 4). We presume that the reduced cellular eGFP content then led to lower metabolic burden and thus enabled slightly higher biomass yields. In all parameters measured, *E. coli* RM214 rhaB⁻ pJOE4056.2_tetA proved to be more robust to the STR-PFR conditions and maintained productive for a longer period than *E. coli* MG1655 rhaB⁻ pJOE4056.2_tetA. Since the only clearly different parameter between the two strains is the maintenance coefficient, we propose that *E. coli* RM214 rhaB⁻ pJOE4056.2_tetA benefits from a small surplus of substrate that can be used to meet the high precursor and ATP demand of heterologous protein synthesis.

The energetic state of cells during cultivations can be assessed by calculating the Adenylate Energy Charge (AEC) from measured nucleotide concentrations (Chapman et al., 1971). Initially, in the well-mixed STR only phase, the concentration of all nucleotides and the AEC was comparable for both strains (Supplementary Fig. S4A). After connection of the PFR, we then simultaneously sampled cells from the STR and the five ports along the primary axis of the PFR to obtain a time-resolved profile of the short-term AEC changes during PFR passage (Fig. 6). As expected during passage through a nutrient starvation zone, the AEC of cells dropped rapidly after leaving the STR and continued to decline towards a plateau. Shortly after PFR connection, the pattern was highly similar for both strains (Fig. 6, upper panel). After 25 h of cultivation under scale-down conditions, the AEC of both strains in the STR and at all sampling ports of the PFR was higher than before (Fig. 6, lower panel). Here, differences between the strains also became apparent as the AEC of *E. coli* MG1655 rhaB⁻ pJOE4056.2_tetA was higher than that of *E. coli* RM214 rhaB⁻ pJOE4056.2_tetA at all sampling points. We then compared the AEC of samples drawn from the primary vessel at different time points to unravel long-term effects of the heterogeneous conditions. Both strains individually showed statistically significant increases in the AEC between time points STR and STR-PFR 25 h (two-tailed t-tests, p <

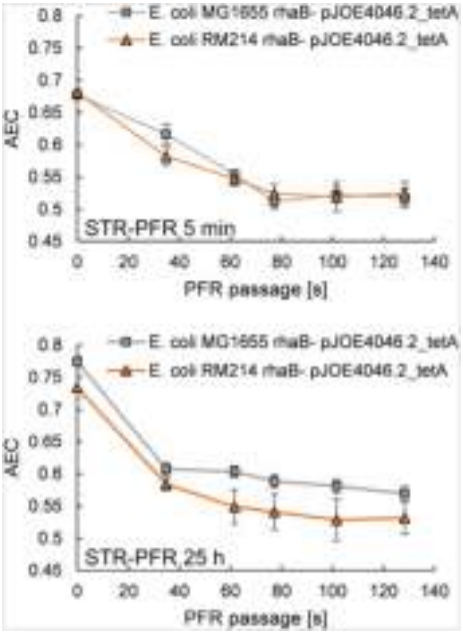


Fig. 6. Adenylate Energy Charge in the STR and during PFR passage. *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA were cultivated in the STR-PFR system (glucose limited chemostat, D = 0.2 h^{-1}). The adenylate energy charge of cultures was determined shortly after PFR connection (STR-PFR 5 min, upper panel) and after five volumetric residence times (STR-PFR 25 h, lower panel). Samples were drawn from the primary reactor (0 s) and the five sampling ports along the axis of the PFR (35 s, 52 s, 77 s, 102 s, 128 s). Error bars indicate SEM (n = 4).

0.05; see Supplementary Table S4B). In fact, the AEC of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA sampled from the primary fermentation vessel (Fig. 6 and 0 s) at time point STR-PFR 25 h was the highest recorded value from all samples indicating that the strain was possibly trying to adapt to the unfavorable conditions. The coincidence with its reduced productivity and slightly increased biomass yield at the late fermentation stages points towards the preservation of cellular energy at the expense of heterologous protein productivity. The data from *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA indicates a similar but less pronounced trend. Comparing the two strains to each other reveals a marginally significant difference ($p = 0.077$; see Supplementary Table S4B) of the AEC values measured in samples from the STR at STR-PFR 25 h which is reflected by the generally slightly lower AEC values of the deletion strain at this time point (Fig. 6, lower panel). It is noteworthy that the total AxP levels of both strains were comparable for all samples and only the distribution among ATP, ADP and AMP varied (Supplementary Fig. S4A).

4. Discussion

In this study, we created a series of deletion strains lacking genes with high add-on maintenance under heterogeneous conditions with repeated starvation. The final strain of the series, *E. coli* RM214 had a significantly lower maintenance coefficient than its parent *E. coli* MG1655 in an STR-PFR scale-down reactor. Moreover, *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA proved to be more robust to the influence of heterogeneities in an exemplary protein production scenario reaching a significantly higher product yield in the STR-PFR phase.

The core concept of our deletion approach was to remove genes that are wastefully expressed under transient starvation conditions. The expected individual contribution of each single gene was very low (Löffler et al., 2016). The only remarkable exception was *fliC* whose expression alone was estimated to cause add-on maintenance of 3.10%, by far exceeding the expected add-on maintenance of 0.55% for the second-in-line *aldA* (Löffler et al., 2016). Multiple other flagellar and chemotaxis genes were also candidates, so the removal of these systems formed a major fraction of the deletions conducted in the creation of *E. coli* RM214. As the goal of this study was to investigate the fundamental usefulness of the whole design approach and each individual deletion had likely little effect, we did not attempt to experimentally assess the individual contributions or potential interactions. The data collected in this study thus serves to prove the general applicability of the approach but is insufficient to test the quantitative reliability of the individual metabolic cost predictions by Löffler et al. (2016). An in-depth study focusing on the effects of subsets or individual deletions would allow judging the predictions and the parallel investigation of fundamental effects and potential interactions. This is particularly interesting as the magnitude of effects observed by us exceeds the sum of all individual contributions from transcriptional and translational metabolic costs as estimated by Löffler et al. (2016). The calculations by Löffler et al. (2016) were thus either very conservative or other effects such as the actual absence of the expressed proteins provided additional secondary benefits. Investigating such secondary benefits could be helpful in refining the target selection among transiently expressed process-irrelevant genes. A potential secondary benefit enjoyed by *E. coli* RM214 could for example arise from the absence of the motility system which could save proton motive force otherwise used for flagellar rotation. On the other hand, we can confidently exclude the possibility that the slightly reduced genome of *E. coli* RM214 had a major impact due to reduced replication cost. The combined size of the deletions in *E. coli* RM214 was only about 50 kb, just slightly more than 1% of the *E. coli* K-12 genome, and the metabolic demand of DNA replication is per se very low (Stouthamer, 1973).

Key findings of our study are the reduced maintenance demand of *E. coli* RM214 compared to its parent strain and the slower product yield decline of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA under STR-PFR conditions. Both differences must be caused by the genotype of *E. coli* RM214 but do these findings relate to each other by a causative link or did we observe correlated phenomena? The overall eGFP transcript levels in the production scenario were fairly stable (Supplementary Information S10), so it appears unlikely that differential expression of eGFP from pJOE4056.2_tetA causes the different yields. The flow cytometry data indicate that microbial individuality may play a role, but without a plausible mechanism we must assume that this is a correlated observation. Instead, we opine that a connecting mechanism can be drawn from the overall balances and the strains' cultivation parameters. Substrate consumed for maintenance demand is, by definition, not available for biomass formation. It is commonly assumed that it is fully converted into terminal products and the energy available from this conversion harnessed by the cells as ATP to meet their maintenance demand (Stouthamer and Bettenhaussen, 1973). Thus, a logical link between maintenance coefficient and protein yield exists: A reduced maintenance coefficient means that more substrate is available for biomass or product formation including potential secondary ATP costs that may arise from a high foreign protein content. A simple estimation allows us to test the quantitative feasibility of a causative relation by comparing the magnitude of the reduced maintenance coefficient to the difference in protein yield: Given that there were no significant differences in Y_{XS}^{true} we can assume that Δm_S is entirely available for the additional production of eGFP in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA.

Using Y_{XS} from STR-PFR 28 h and an assumed protein content of roughly 65% (Taymaz-Nikerel et al., 2010) the difference in maintenance demand could sustain an additional eGFP production rate of no more than $\Delta q_{p,ms} = 9.3\ mg_{eGFP}/g_{CDW}/h$. The experimental difference of $\Delta q_{p,ms} = 7\ mg_{eGFP}/g_{CDW}/h$ falls well within that range. A causative relation between the two observations is thus quantitatively feasible, and in our opinion likely. In this case up to 82% of the saved substrate due to lower m_s could have been used for the formation of additional eGFP in *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA.

The connection between maintenance demand, energy availability and eGFP production is also supported by the AEC data collected. We found a declining AEC during PFR passage for both *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA and *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA at all time points, which is similar to the pattern observed in a preceding study with non-producing *E. coli* K-12 (Löffler et al., 2016). However, it is important to note that we measured lower AEC values in the STR and a steeper decline in the PFR, putatively due to heterologous protein production. In the heterogeneous fermentation phase, when productivity declined, we measured increased AEC values, especially in samples of the less productive *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA (Supplementary Information S4). The AEC is a measurement of the energetic state of cells and usually tightly balanced in a range between 0.7 and 0.9 (Chapman et al., 1971). The activity of many cellular processes is connected to the AEC and a lower AEC is associated with the activation of catabolic enzymes to meet cellular energy demands (Atkinson, 1968). Substrate depletion generally causes a reduction of the AEC (Chapman and Atkinson, 1977). Conversely, reduced AEC values have been reported in conditions when cells experienced high anabolic demand or high secondary metabolic costs, for instance caused by cultivation at their maximum specific growth rate, or induction of motility (Lieder et al., 2015; Martinez-Garcia et al., 2014). Heterologous protein induction is known to cause increased ATP maintenance demand (Weber et al., 2002). We thus propose that the AEC values measured from samples drawn from the primary reactor at different time points of the fermentations can be explained by the ATP demands associated with eGFP productivity. It appears likely that the generally lower AEC measured in this study compared to data from non-producing *E. coli* K-12 cultivated under similar conditions is caused by the production of eGFP. The significant increases of the AEC values of both strains towards the end of the fermentations are then a consequence of their diminishing eGFP productivity. This also explains the more pronounced AEC increase and concomitant eGFP yield decrease of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA compared to the deletion strain. The question then arises to what extent the lower maintenance coefficient of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA under scale-down conditions influences the AEC values. From data collected in the maintenance study (Fig. 3) and the eGFP production fermentations we can roughly estimate the ATP demand for eGFP production of both strains at time point STR-PFR 25 h (Supplementary Data S8). About 13% of the total ATP demand of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and 22% of the total ATP demand of *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA can be attributed to eGFP production. Despite its lower maintenance coefficient the combined ATP demand for maintenance plus eGFP production of the highly productive *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA then still exceeds the respective values of *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA which is reflected by its lower AEC at this time point.

A secondary observation made in this study was that loss of productivity in the STR-PFR condition was accompanied by a decline in the proportion of highly productive cells (Fig. 4). Microbial population heterogeneity is a subject of intense research (Binder et al., 2017) and our data provides no clear explanation why this shift occurs. Two things should be noted: First, the population heterogeneity for both *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA and *E. coli* RM214 rhaB$^-$ pJOE4056.2_tetA is of the bimodal kind (Supplementary Fig. S3) and the fractions of producing cells in the homogeneous STR cultivation phase are practically identical for both strains. Second, once the PFR is activated, we saw a decrease in the fraction of highly productive cells in all fermentations (Fig. 4, Supplementary Fig. S3) but the decrease was faster and more consistent for *E. coli* MG1655 rhaB$^-$ pJOE4056.2_tetA. The overall level of population heterogeneity is generally high since only slightly more than half of all cells are strongly accumulating eGFP. We presume this is caused by the interplay of our expression system and the fermentation conditions. The regulation of rhamnose catabolism is autocatalytic and thus bimodality might be caused by a similar mechanism as in the case of expression from the arabinose promoter P_{BAD} (Khlebnikov et al., 2000). However, transcript measurements by RT-qPCR did not lead to a concise pattern that would explain both the differences between the two strains and the declining eGFP yield of each individual strain over the course of the heterogeneous fermentation phase. Given the fairly stable expression of eGFP and the continuous selective advantage provided by *tetA*, it also appears unlikely that plasmid loss or mutations were the underlying cause. Moreover, the general stability of eGFP expression from pJOE4056.2 has been determined to be perfect for over 50 generations in earlier studies (Wegerer et al., 2008).

The deletion approach in this study differs from previous works because target selection was based on existing expression data and limited to candidates that imposed a high metabolic burden but were irrelevant under the specified conditions (Valgepea et al., 2015). Large-scale genomic deletions, the contrary approach, have been conducted before in *E. coli* K-12, for instance as part of the construction of the MDS strains (Posfai et al., 2006). These strains had little benefits in standard protein production scenarios over their wild-type parent and were even inferior in basic process parameters, potentially caused by disrupted regulation (Karcagi et al., 2016; Sharma et al., 2007a, 2007b). It needs to be emphasized that our deletion strategy only provided advantages in the specified conditions of a heterogeneous bioprocess with transient starvation as *E. coli* RM214 had no benefit compared to *E. coli* MG1655 under well-mixed conditions. In this regard, it also appears clear that the deletion targets chosen by us cannot be directly transferred to other hosts or conditions since the naturally evolved regulation might be divergent. This is exemplified by an interesting comparison of our results to existing data from *Pseudomonas putida.* The exposure of *P. putida* KT2440 to heterogeneous STR-PFR conditions led to an increased and potentially wasteful expression of *fliC* similarly as in the case of *E. coli* (Ankenbauer et al., 2020). However, the deletion of the flagellar apparatus in *P. putida* EM329 led to significant improvements of basic fermentation and protein production parameters already under well mixed conditions (Lieder et al., 2015; Martinez-Garcia et al., 2014). This demonstrates not only the diverging regulation of motility between different microbes, but also implies that a strain like *P. putida* EM329 might unintentionally display additional beneficial traits in heterogeneous fermentations with starvation zones.

A premise of our study was to work orthogonally to global cellular regulation. The general stress response, stringent response and SOS responses were not modified as they might provide important functions for cellular adaptation and some expression systems depend on intracellular signaling molecules of global regulatory circuits. Examples include not only the CRP-cAMP dependent system used in this study but also novel adaptive expression systems that autoregulate cellular stress (Lo et al., 2016). Interestingly, in a former study, the rapid inactivation of *rpoS* in homogeneous chemostat cultivations was reported, which pointed to a large selective advantage of mutants (Notley-McRobb et al., 2002). In contrast, we did not find any mutations in genes involved in the stringent response or the general stress response for neither *E. coli* MG1655 nor *E. coli* RM214 at any growth rate. We conclude that under heterogeneous conditions the selective pressure on inactivating global regulatory programs is either very low or their activation may even be favorable for cellular viability which affirms our neutral approach to cellular regulation. However, this does not mean that modulating cellular regulation could not be beneficial for process or production parameters. Recently, several *E. coli* knock out strains lacking

hand-picked genes that are connected to post-induction stress responses were presented (Sharma et al., 2020). These strains have advantageous traits for protein production which could be integrated into *E. coli* RM214.

Our study design focused on the influence of starvation zones on microbial culture performance. The carbon limited STR contained a substrate-limited growth and production zone representing the bulk of large-scale fermentations. The PFR served to introduce a transient starvation stimulus representative of repeated passages through a hunger zone as predicted to occur in large-scale reactors (Haringa et al., 2017; Kuschel and Takors, 2020). Since our experimental setup was specifically chosen for the study of transient starvation, it does not capture the effects of other heterogeneities, in particular transient substrate excess. It is well-known that close to the feeding point, substrate excess and concomitant oxygen limitation dominate the environment in large-scale fed-batch processes. *E. coli* typically reacts to such conditions with the production of solvents or small organic acids caused by overflow metabolism or anaerobic fermentation (Lara et al., 2009). The formation of byproducts then results in process performance losses even if reuptake in zones with lower nutrient concentration is possible (Enfors et al., 2001; Neubauer et al., 1995b). Since our deletion approach was only aimed at reducing the additional metabolic costs of transient starvation, *E. coli* RM214 probably responds to glucose excess like other *E. coli* K-12 strains. In principle, copying the design approach to construct an *E. coli* strain with reduced additional maintenance in excess zones appears to be feasible as the transcriptional response of *E. coli* MG1655 to glucose excess is large and involves many potentially process-irrelevant genes (Veit et al., 2007). However, the resulting genetic modifications would not reduce process performance loss from byproduct formation which is likely the dominating issue in substrate excess conditions. Various strategies to alleviate byproduct metabolism have been developed by other research groups, such as the use of alternative substrate transporters, knock-outs or the expression of recombinant *Vitreoscilla* hemoglobin (Anda et al., 2006; Eiteman and Altman, 2006; Pablos et al., 2014). Given that our strain design approach avoids modifications to global regulation or central carbon metabolism, we are confident that it is compatible with these existing strategies and their combination could result in chassis strains for generally robust scale-up. A limitation of our study originates from the focus on the model protein eGFP. However, recombinant protein production is frequently limited by the availability of cellular precursors and ATP, so it is not far-fetched to expect similar effects with other protein products (Glick, 1995; Heyland et al., 2011). The reduced maintenance coefficient of *E. coli* RM214 should also be helpful to produce molecules formed in ATP-intensive pathways such as terpenoids (Li and Wang, 2016; Ward et al., 2018). Potential advantages could also occur when the accumulation of toxic products causes increased ATP demand for product export, membrane maintenance or pH homeostasis (Sun et al., 2011; Tsukagoshi and Aono, 2000). On the other hand, it may be less helpful when the formation of a small molecule product is connected to net ATP synthesis. Glycolytic flux depends on the ATP requirements of cells and in such cases enforced ATP wasting can even increase the production rate (Boecker et al., 2019; Koebmann et al., 2002).

With the increasing knowledge about cellular metabolism and its interplay with the heterogeneous conditions of large-scale processes new possibilities to improve process performance arise. In a recent review Wehrs et al. (2019) emphasized that strains should be engineered specifically for the demands of large-scale production (Wehrs et al., 2019). In this context, our series of deletion strains is the first step towards host strains robust against the repeated exposure to starvation zones.

5. Conclusion

Large-scale fermentations often suffer from process performance loss due to heterogeneous environments. *E. coli RM214* was engineered to obtain a deletion strain with reduced maintenance and superior production properties in fermentations with starvation zones. Our study is the first that aimed to improve a microbe by repeated genomic deletions for enhanced robustness towards heterogeneous conditions. The exemplified application of *E. coli* RM214 for eGFP production demonstrates the cellular capacity to exploit the maintenance advantage for preventing non-wanted performance loss in heterogeneously mixed industrial production scenarios. Although only showcased for eGFP, the strain offers the capacity to serve as a platform for a variety of different products. Notably, this complements classical scale-up engineering and should be a highly valuable tool to prevent non-wanted performance of essential Titer-Rate-Yield values under industrial production conditions.

Funding

This research was funded from discretionary funds provided by the University of Stuttgart.

CRediT authorship contribution statement

Martin Ziegler: Conceptualization, Methodology, Investigation, Writing – original draft, Visualization. **Julia Zieringer:** Software, Formal analysis. **Clarissa-Laura Döring:** Methodology, Investigation. **Liv Paul:** Investigation. **Christoph Schaal:** Investigation. **Ralf Takors:** Conceptualization, Writing – review & editing, Supervision.

Declaration of competing interest

Authors Martin Ziegler and Ralf Takors declare that findings of this manuscript contents are part of a patent application submitted by the privately owned company Manus Bio (Massachusetts, USA). The authors Martin Ziegler and Ralf Takors are listed as inventors.

Acknowledgements

The authors acknowledge the support of the Computational Biology Group at IBVT Stuttgart: Richard Schäfer and Prof. Björn Voβ for providing access to a Galaxy Server.

The authors acknowledge the support of Dr. rer. nat. habil. Jürgen Dippon for assistance with statistical analysis.

The authors acknowledge the support of Florian Hiering and apl. Prof. Martin Siemann-Herzberg for assistance with the purification and quantification of eGFP.

The authors acknowledge the support of Andreas Freund for technical assistance.

The authors would like to thank the staff of the Court laboratory at the National Cancer Institute in Frederick, MD USA for providing the strains *T-SACK* and *E. coli* DH10B pSIM5.

The authors would like to thank apl. Prof. Martin Siemann-Herzberg and Dr. Josef Altenbuchner for providing the strain *E. coli* BW3110 pJOE4056.2.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymben.2021.05.011.

References

Anda, R. de, Lara, A.R., Hernández, V., Hernández-Montalvo, V., Gosset, G., Bolívar, F., Ramírez, O.T., 2006. Replacement of the glucose phosphotransferase transport system by galactose permease reduces acetate accumulation and improves process performance of Escherichia coli for recombinant protein production without impairment of growth rate. Metab. Eng. 8, 281–290. https://doi.org/10.1016/j.ymben.2006.01.002.

Ankenbauer, A., Schäfer, R.A., Viegas, S.C., Pobre, V., Voß, B., Arraiano, C.M., Takors, R., 2020. Pseudomonas putida KT2440 is naturally endowed to withstand industrial-scale stress conditions. Microbial biotechnology 13, 1145–1161. https://doi.org/10.1111/1751-7915.13571.

Atkinson, D.E., 1968. The energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers. Biochemistry 7, 4030–4034. https://doi.org/10.1021/bi00851a033.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2 https://doi.org/10.1038/msb4100050.

Baeshen, M.N., Al-Hejin, A.M., Bora, R.S., Ahmed, M.M.M., Ramadan, H.A.I., Saini, K.S., Baeshen, N.A., Redwan, E.M., 2015. Production of biopharmaceuticals in E. coli: current scenario and future perspectives. J. Microbiol. Biotechnol. 25, 953–962. https://doi.org/10.4014/jmb.1412.12079.

Baeshen, N.A., Baeshen, M.N., Sheikh, A., Bora, R.S., Ahmed, M.M.M., Ramadan, H.A.I., Saini, K.S., Redwan, E.M., 2014. Cell factories for insulin production. Microb. Cell Factories 13, 141. https://doi.org/10.1186/s12934-014-0141-0.

Binder, D., Drepper, T., Jaeger, K.-E., Delvigne, F., Wiechert, W., Kohlheyer, D., Grünberger, A., 2017. Homogenizing bacterial cell factories: analysis and engineering of phenotypic heterogeneity. Metab. Eng. 42, 145–156. https://doi.org/10.1016/j.ymben.2017.06.009.

Boecker, S., Zahoor, A., Schramm, T., Link, H., Klamt, S., 2019. Broadening the scope of enforced ATP wasting as a tool for metabolic engineering in Escherichia coli. Biotechnol. J. 14, e1800438 https://doi.org/10.1002/biot.201800438.

Bylund, F., Castan, A., Mikkola, R., Veide, A., Larsson, G., 2000. Influence of scale-up on the quality of recombinant human growth hormone. Biotechnol. Bioeng. 69, 119–128. https://doi.org/10.1002/(SICI)1097-0290(20000720)69:2<119:AID-BIT1>3.0.CO;2-9.

Bylund, F., Collet, E., Enfors, S.-O., Larsson, G., 1998. Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. Bioprocess Eng. 18, 171. https://doi.org/10.1007/s004490050427.

Bylund, F., Guillard, F., Enfors, S.-O., Trägårdh, C., Larsson, G., 1999. Scale down of recombinant protein production: a comparative study of scaling performance. Bioprocess Eng. 20, 377. https://doi.org/10.1007/s004490050606.

Chapman, A.G., Atkinson, D.E., 1977. Adenine nucleotide concentrations and turnover rates. Their correlation with biological activity in bacteria and yeast. In: Rose, A.H., Tempest, D.W. (Eds.), Advances in Microbial Physiology, vol. 15. Academic, London, pp. 253–306.

Chapman, A.G., Fall, L., Atkinson, D.E., 1971. Adenylate energy charge in Escherichia coli during growth and starvation. J. Bacteriol. 108, 1072–1086.

Cortés, J.T., Flores, N., Bolívar, F., Lara, A.R., Ramírez, O.T., 2016. Physiological effects of pH gradients on Escherichia coli during plasmid DNA production. Biotechnol. Bioeng. 113, 598–611. https://doi.org/10.1002/bit.25817.

Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403. https://doi.org/10.1101/gr.2289704.

Datta, S., Costantino, N., Court, D.L., 2006. A set of recombineering plasmids for gram-negative bacteria. Gene 379, 109–115. https://doi.org/10.1016/j.gene.2006.04.018.

Delvigne, F., Boxus, M., Ingels, S., Thonart, P., 2009. Bioreactor mixing efficiency modulates the activity of a prpoS:GFP reporter gene in E. coli. Microb. Cell Factories 8, 15. https://doi.org/10.1186/1475-2859-8-15.

Delvigne, F., Destain, J., Thonart, P., 2006. A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models. Biochem. Eng. J. 28, 256–268. https://doi.org/10.1016/j.bej.2005.11.009.

Delvigne, F., Takors, R., Mudde, R., van Gulik, W., Noorman, H., 2017. Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. Microbial biotechnology 10, 1267–1274. https://doi.org/10.1111/1751-7915.12803.

Eiteman, M.A., Altman, E., 2006. Overcoming acetate in Escherichia coli recombinant protein fermentations. Trends Biotechnol. 24, 530–536. https://doi.org/10.1016/j.tibtech.2006.09.001.

Enfors, S.-O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O'Beirne, D., Noisommit-Rizzi, N., Reuss, M., Boone, L., Hewitt, C., McFarlane, C., Nienow, A., Kovacs, T., Trägårdh, C., Fuchs, L., Revstedt, J., Friberg, P.C., Hjertager, B., Blomsten, G., Skogman, H., Hjort, S., Hoeks, F., Lin, H.-Y., Neubauer, P., van der Lans, R., Luyben, K., Vrabel, P., Manelius, Å., 2001. Physiological responses to mixing in large scale bioreactors. J. Biotechnol. 85, 175–185. https://doi.org/10.1016/S0168-1656(00)00365-5.

George, S., Larsson, G., Enfors, S.-O., 1993. A scale-down two-compartment reactor with controlled substrate oscillations: metabolic response of Saccharomyces cerevisiae. Bioprocess Eng. 9, 249–257. https://doi.org/10.1007/BF01061530.

Glick, B.R., 1995. Metabolic load and heterologous gene expression. Biotechnol. Adv. 13, 247–261. https://doi.org/10.1016/0734-9750(95)00004-A.

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166, 557–580. https://doi.org/10.1016/S0022-2836(83)80284-8.

Haringa, C., Deshmukh, A.T., Mudde, R.F., Noorman, H.J., 2017. Euler-Lagrange analysis towards representative down-scaling of a 22 m 3 aerobic S. cerevisiae fermentation. Chem. Eng. Sci. 170, 653–669. https://doi.org/10.1016/j.ces.2017.01.014.

Heyland, J., Blank, L.M., Schmid, A., 2011. Quantification of metabolic limitations during recombinant protein production in Escherichia coli. J. Biotechnol. 155, 178–184. https://doi.org/10.1016/j.jbiotec.2011.06.016.

Jonge, L.P. de, Buijs, N.A.A., Pierick, A. ten, Deshmukh, A., Zhao, Z., Kiel, J.A.K.W., Heijnen, J.J., van Gulik, W.M., 2011. Scale-down of penicillin production in Penicillium chrysogenum. Biotechnol. J. 6, 944–958. https://doi.org/10.1002/biot.201000409.

Junne, S., Klingner, A., Kabisch, J., Schweder, T., Neubauer, P., 2011. A two-compartment bioreactor system made of commercial parts for bioprocess scale-down studies: impact of oscillations on Bacillus subtilis fed-batch cultivations. Biotechnol. J. 6, 1009–1017. https://doi.org/10.1002/biot.201100293.

Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovács, K., Méhi, O., Balikó, G., Szappanos, B., Györfy, Z., Fehér, T., Bogos, B., Blattner, F.R., Pál, C., Pósfai, G., Papp, B., 2016. Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining. Mol. Biol. Evol. 33, 1257–1269. https://doi.org/10.1093/molbev/msw009.

Khlebnikov, A., Risa, Ø., Skaug, T., Carrier, T.A., Keasling, J.D., 2000. Regulatable Arabinose-inducible gene expression system with consistent control in all cells of a culture. J. Bacteriol. 182, 7029–7034. https://doi.org/10.1128/JB.182.24.7029-7034.2000.

Koebmann, B.J., Westerhoff, H.V., Snoep, J.L., Nilsson, D., Jensen, P.R., 2002. The glycolytic flux in Escherichia coli is controlled by the demand for ATP. J. Bacteriol. 184, 3909–3916. https://doi.org/10.1128/JB.184.14.3909–3916.2002.

Kurokawa, M., Seno, S., Matsuda, H., Ying, B.-W., 2016. Correlation between genome reduction and bacterial growth. DNA Res. : an international journal for rapid publication of reports on genes and genomes 23, 517–525. https://doi.org/10.1093/dnares/dsw035.

Kuschel, M., Siebler, F., Takors, R., 2017. Lagrangian Trajectories to Predict the Formation of Population Heterogeneity in Large-Scale Bioreactors, vol. 4. Bioengineering, Basel, Switzerland. https://doi.org/10.3390/bioengineering4020027.

Kuschel, M., Takors, R., 2020. Simulated oxygen and glucose gradients as a prerequisite for predicting industrial scale performance a priori. Biotechnol. Bioeng. 117, 2760–2770. https://doi.org/10.1002/bit.27457.

Lara, A.R., Galindo, E., Ramírez, O.T., Palomares, L.A., 2006a. Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. MB 34, 355–382. https://doi.org/10.1385/MB:34:3:355.

Lara, A.R., Leal, L., Flores, N., Gosset, G., Bolívar, F., Ramírez, O.T., 2006b. Transcriptional and metabolic response of recombinant Escherichia coli to spatial dissolved oxygen tension gradients simulated in a scale-down system. Biotechnol. Bioeng. 93, 372–385. https://doi.org/10.1002/bit.20704.

Lara, A.R., Taymaz-Nikerel, H., Mashego, M.R., van Gulik, W.M., Heijnen, J.J., Ramírez, O.T., van Winden, W.A., 2009. Fast dynamic response of the fermentative metabolism of Escherichia coli to aerobic and anaerobic glucose pulses. Biotechnol. Bioeng. 104, 1153–1161. https://doi.org/10.1002/bit.22503.

Larsson, G., Enfors, S.-O., 1988. Studies of insufficient mixing in bioreactors: effects of limiting oxygen concentrations and short term oxygen starvation on Penicillium chrysogenum. Bioprocess Eng. 3, 123–127. https://doi.org/10.1007/BF00373475.

Li, X.-T., Thomason, L.C., Sawitzke, J.A., Costantino, N., Court, D.L., 2013. Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli. Nucleic Acids Res. 41, e204 https://doi.org/10.1093/nar/gkt1075.

Li, Y., Wang, G., 2016. Strategies of isoprenoids production in engineered bacteria. J. Appl. Microbiol. 121, 932–940. https://doi.org/10.1111/jam.13237.

Lieder, S., Nikel, P.I., Lorenzo, V. de, Takors, R., 2015. Genome reduction boosts heterologous gene expression in Pseudomonas putida. Microb. Cell Factories 14, 23. https://doi.org/10.1186/s12934-015-0207-7.

Lo, T.-M., Chng, S.H., Teo, W.S., Cho, H.-S., Chang, M.W., 2016. A two-layer gene circuit for decoupling cell growth from metabolite production. Cell Syst 3, 133–143. https://doi.org/10.1016/j.cels.2016.07.012.

Löffler, M., Simen, J.D., Jager, G., Schaferhoff, K., Freund, A., Takors, R., 2016. Engineering E. coli for large-scale production - strategies considering ATP expenses and transcriptional responses. Metab. Eng. 38, 73–85. https://doi.org/10.1016/j.ymben.2016.06.008.

Martinez-Garcia, E., Nikel, P.I., Chavarria, M., Lorenzo, V. de, 2014. The metabolic cost of flagellar motion in Pseudomonas putida KT2440. Environ. Microbiol. 16, 291–303. https://doi.org/10.1111/1462-2920.12309.

Michalowski, A., Siemann-Herzberg, M., Takors, R., 2017. Escherichia coli HGT: engineered for high glucose throughput even under slowly growing or resting conditions. Metab. Eng. 40, 93–103. https://doi.org/10.1016/j.ymben.2017.01.005.

Murphy, K.C., 2016. λ recombination and recombineering. EcoSal Plus 7. https://doi.org/10.1128/ecosalplus.ESP-0011-2015.

Neubauer, P., Åhman, M., Törnkvist, M., Larsson, G., Enfors, S.-O., 1995a. Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of Escherichia coli. J. Biotechnol. 43, 195–204. https://doi.org/10.1016/0168-1656(95)00130-1.

Neubauer, P., Häggström, L., Enfors, S.O., 1995b. Influence of substrate oscillations on acetate formation and growth yield in Escherichia coli glucose limited fed-batch cultivations. Biotechnol. Bioeng. 47, 139–146. https://doi.org/10.1002/bit.260470204.

Neubauer, P., Junne, S., 2010. Scale-down simulators for metabolic analysis of large-scale bioprocesses. Curr. Opin. Biotechnol. 21, 114–121. https://doi.org/10.1016/j.copbio.2010.02.001.

Nieß, A., Löffler, M., Simen, J.D., Takors, R., 2017. Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of E. coli cultures in large-scale bioreactors: experimental evidence and mathematical model. Front. Microbiol. 8, 1195. https://doi.org/10.3389/fmicb.2017.01195.

Noorman, H., 2011. An industrial perspective on bioreactor scale-down: what we can learn from combined large-scale bioprocess and model fluid studies. Biotechnol. J. 6, 934–943. https://doi.org/10.1002/biot.201000406.

Notley-McRobb, L., King, T., Ferenci, T., 2002. rpoS mutations and loss of general stress resistance in Escherichia coli populations as a consequence of conflict between competing stress responses. J. Bacteriol. 184, 806–811. https://doi.org/10.1128/JB.184.3.806-811.2002.

Olughu, W., Deepika, G., Hewitt, C., Rielly, C., 2019. Insight into the large-scale upstream fermentation environment using scaled-down models. J. Chem. Technol. Biotechnol. 94, 647–657. https://doi.org/10.1002/jctb.5804.

Olughu, W., Nienow, A., Hewitt, C., Rielly, C., 2020. Scale-down studies for the scale-up of a recombinant Corynebacterium glutamicum fed-batch fermentation: loss of homogeneity leads to lower levels of cadaverine production. J. Chem. Technol. Biotechnol. 95, 675–685. https://doi.org/10.1002/jctb.6248.

Pablos, T.E., Sigala, J.C., Le Borgne, S., Lara, A.R., 2014. Aerobic expression of Vitreoscilla hemoglobin efficiently reduces overflow metabolism in Escherichia coli. Biotechnol. J. 9, 791–799. https://doi.org/10.1002/biot.201300388.

Pirt, S.J., 1965. The maintenance energy of bacteria in growing cultures. Proc. Roy. Soc. Lond. B Biol. Sci. 163, 224–231. https://doi.org/10.1098/rspb.1965.0069.

Posfai, G., Plunkett, G.3., Feher, T., Frisch, D., Keil, G.M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S.S., Arruda, M. de, Burland, V., Harcum, S.W., Blattner, F.R., 2006. Emergent properties of reduced-genome Escherichia coli. Science 312, 1044–1046. https://doi.org/10.1126/science.1126439.

Sharma, A.K., Shukla, E., Janoti, D.S., Mukherjee, K.J., Shiloach, J., 2020. A novel knock out strategy to enhance recombinant protein expression in Escherichia coli. Microb. Cell Factories 19. https://doi.org/10.1186/s12934-020-01407-z.

Sharma, S.S., Blattner, F.R., Harcum, S.W., 2007a. Recombinant protein production in an Escherichia coli reduced genome strain. Metab. Eng. 9, 133–141. https://doi.org/10.1016/j.ymben.2006.10.002.

Sharma, S.S., Campbell, J.W., Frisch, D., Blattner, F.R., Harcum, S.W., 2007b. Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome Escherichia coli strains. Biotechnol. Bioeng. 98, 1056–1070. https://doi.org/10.1002/bit.21491.

Simen, J.D., Löffler, M., Jäger, G., Schäferhoff, K., Freund, A., Matthes, J., Müller, J., Takors, R., 2017. Transcriptional response of Escherichia coli to ammonia and glucose fluctuations. Microbial biotechnology 10, 858–872. https://doi.org/10.1111/1751-7915.12713.

Stouthamer, A.H., 1973. A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie Leeuwenhoek 39, 545–565. https://doi.org/10.1007/BF02578899.

Stouthamer, A.H., Bettenhaussen, C., 1973. Utilization of energy for growth and maintenance in continuous and batch cultures of microorganisms. Biochim. Biophys. Acta Rev. Bioenerg. 301, 53–70. https://doi.org/10.1016/0304-4173(73)90012-8.

Sun, Y., Fukamachi, T., Saito, H., Kobayashi, H., 2011. ATP requirement for acidic resistance in Escherichia coli. J. Bacteriol. 193, 3072–3077. https://doi.org/10.1128/JB.00091-11.

Sunya, S., Gorret, N., Delvigne, F., Uribelarrea, J.-L., Molina-Jouve, C., 2012. Real-time monitoring of metabolic shift and transcriptional induction of yciG:luxCDABE E. coli reporter strain to a glucose pulse of different concentrations. J. Biotechnol. 157, 379–390. https://doi.org/10.1016/j.jbiotec.2011.12.009.

Takors, R., 2012. Scale-up of microbial processes: impacts, tools and open questions. J. Biotechnol. 160, 3–9. https://doi.org/10.1016/j.jbiotec.2011.12.010.

Taymaz-Nikerel, H., Borujeni, A.E., Verheijen, P.J.T., Heijnen, J.J., van Gulik, W.M., 2010. Genome-derived minimal metabolic models for Escherichia coli MG1655 with estimated in vivo respiratory ATP stoichiometry. Biotechnol. Bioeng. 107, 369–381. https://doi.org/10.1002/bit.22802.

Tomasek, K., Bergmiller, T., Guet, C.C., 2018. Lack of cations in flow cytometry buffers affect fluorescence signals by reducing membrane stability and viability of Escherichia coli strains. J. Biotechnol. 268, 40–52. https://doi.org/10.1016/j.jbiotec.2018.01.008.

Tsukagoshi, N., Aono, R., 2000. Entry into and release of solvents by Escherichia coli in an organic-aqueous two-liquid-phase system and substrate specificity of the AcrAB-TolC solvent-extruding pump. J. Bacteriol. 182, 4803–4810. https://doi.org/10.1128/jb.182.17.4803-4810.2000.

Valgepea, K., Peebo, K., Adamberg, K., Vilu, R., 2015. Lean-proteome strains - next step in metabolic engineering. Frontiers in bioengineering and biotechnology 3, 11. https://doi.org/10.3389/fbioe.2015.00011.

Vasilakou, E., Loosdrecht, van, Mark, C.M., Wahl, S.A., 2020. Escherichia coli metabolism under short-term repetitive substrate dynamics: adaptation and trade-offs. Microb. Cell Factories 19, 116. https://doi.org/10.1186/s12934-020-01379-0.

Veit, A., Polen, T., Wendisch, V.F., 2007. Global gene expression analysis of glucose overflow metabolism in Escherichia coli and reduction of aerobic acetate formation. Appl. Microbiol. Biotechnol. 74, 406–421. https://doi.org/10.1007/s00253-006-0680-3.

Ward, V.C.A., Chatzivasileiou, A.O., Stephanopoulos, G., 2018. Metabolic engineering of Escherichia coli for the production of isoprenoids. FEMS Microbiol. Lett. 365 https://doi.org/10.1093/femsle/fny079.

Weber, J., Hoffmann, F., Rinas, U., 2002. Metabolic adaptation of Escherichia coli during temperature-induced recombinant protein production: 2. Redirection of metabolic fluxes. Biotechnol. Bioeng. 80, 320–330. https://doi.org/10.1002/bit.10380.

Wegerer, A., Sun, T., Altenbuchner, J., 2008. Optimization of an E. coli L-rhamnose-inducible expression vector: test of various genetic module combinations. BMC Biotechnol. 8, 2. https://doi.org/10.1186/1472-6750-8-2.

Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T.R., Mukhopadhyay, A., 2019. Engineering robust production microbes for large-scale cultivation. Trends Microbiol. 27, 524–537. https://doi.org/10.1016/j.tim.2019.01.006.

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, e1005595 https://doi.org/10.1371/journal.pcbi.1005595.

Wilms, B., Hauck, A., Reuss, M., Syldatk, C., Mattes, R., Siemann, M., Altenbuchner, J., 2001. High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter. Biotechnol. Bioeng. 73, 95–103. https://doi.org/10.1002/bit.1041.

Ziegler, M., Zieringer, J., Takors, R., 2020. Transcriptional profiling of the stringent response mutant strain E. coli SR reveals enhanced robustness to large-scale conditions. Microbial biotechnology. https://doi.org/10.1111/1751-7915.13738.

Zieringer, J., Wild, M., Takors, R., 2020. Data-driven in silico prediction of regulation heterogeneity and ATP demands of Escherichia coli in large-scale bioreactors. Biotechnol. Bioeng. https://doi.org/10.1002/bit.27568.

microbial biotechnology

Open Access

Transcriptional profiling of the stringent response mutant strain *E. coli* SR reveals enhanced robustness to large-scale conditions

Martin Ziegler, Julia Zieringer and Ralf Takors*
Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, Germany.

Summary

In large-scale fed-batch production processes, microbes are exposed to heterogeneous substrate availability caused by long mixing times. *Escherichia coli*, the most common industrial host for recombinant protein production, reacts by recurring accumulation of the alarmone ppGpp and energetically wasteful transcriptional strategies. Here, we compare the regulatory responses of the stringent response mutant strain *E. coli SR* and its parent strain *E. coli* MG1655 to repeated nutrient starvation in a two-compartment scale-down reactor. Our data show that *E. coli* SR can withstand these stress conditions without a ppGpp-mediated stress response maintaining fully functional ammonium uptake and biomass formation. Furthermore, *E. coli* SR exhibited a substantially reduced short-term transcriptional response compared to *E. coli* MG1655 (less than half as many differentially expressed genes). *E. coli* SR proceeded adaptation via more general SOS response pathways by initiating negative regulation of transcription, translation and cell division. Our results show that locally induced stress responses propagating through the bioreactor do not result in cyclical induction and repression of genes in *E. coli* SR, but in a reduced and coordinated response, which makes it potentially suitable for large-scale production processes.

Introduction

Heterogeneities in large-scale fed-batch bioprocesses have long been recognized as a cause for process performance loss at industrial scale compared to homogeneous processes at laboratory scale (Bylund *et al.*, 1998). Due to physical, economical and engineering constraints, the generation of gradients in large-scale reactors is inevitable. Hydrostatic pressure influences the solubility and transfer of gasses, and the mixing time of large reactors can be orders of magnitude higher than that of laboratory reactors producing strong measurable chemical gradients (Larsson *et al.*, 1996; Enfors *et al.*, 2001; Junker, 2004; Delvigne *et al.*, 2006). Common consequences of spatial heterogeneities are loss of productivity, reduced biomass yield, increased byproduct formation and genetic or plasmid instability (Hopkins *et al.*, 1987; George *et al.*, 1993; Neubauer *et al.*, 1995b; Bylund *et al.*, 1998; Bylund *et al.*, 2000; Jonge *et al.*, 2011). Reduced process performance is not limited to a single species but can be observed for many industrial workhorse organisms like *Escherichia coli*, *Saccharomyces cerevisiae, Penicillium chrysogenum* and *Bacillus subtilis* (George *et al.*, 1993; Jonge *et al.*, 2011; Junne *et al.*, 2011; Larsson and Enfors, 1988).

Due to the enormous costs associated with using and maintaining large-scale equipment, few experiments in the context of academic research have been performed in industrial scale bioreactors (Bylund *et al.*, 1999; Bylund *et al.*, 2000; Enfors *et al.*, 2001). In consequence, researchers have relied on the use of computational fluid dynamics (CFD) to simulate reactor flow fields and on scale-down reactors to experimentally investigate selected scenarios (Kelly, 2008; Takors, 2012). Various designs of scale-down reactors exist and have been extensively reviewed elsewhere (Delvigne *et al.*, 2017; Delvigne *et al.*, 2006; Neubauer and Junne, 2010). One of the commonly used scale-down reactors follows a multi-compartment approach: A primary stirred tank reactor (STR) is coupled to a secondary plug flow reactor (PFR). The STR is operated as a well-mixed compartment under standard limited growth conditions and the PFR simulates a feeding, starvation or anaerobic zone providing the stimulus to be investigated (Lara *et al.*, 2006).

Many studies have focused on experimentally simulating the zone close to the feeding point which is usually characterized by substrate excess and potentially oxygen limitation (Enfors *et al.*, 2001; Lara *et al.*, 2009;

Received 12 October, 2020; revised 8 December, 2020; accepted 8 December, 2020.
*For correspondence. E-mail ralf.takors@ibvt.uni-stuttgart.de; Tel. +49 711 685-64535; Fax +49 711 685-55164.
Funding information No funding information provided.
Microbial Biotechnology (2020) **0**(0), 1–18
doi:10.1111/1751-7915.13738

Junne *et al.*, 2011). For a variety of hosts, common observations in this scenario include the formation of small organic acids and solvents as overflow metabolites or as anaerobic fermentation products (George *et al.*, 1993; Neubauer *et al.*, 1995b). Ultimately, byproduct formation may lead to process performance loss even if reuptake of byproducts occurs in the well-mixed limited growth zone (Enfors *et al.*, 2001).

Occasionally, starvation zones have attracted attention as well (Neubauer *et al.*, 1995a; Neubauer *et al.*, 1995b). From CFD simulation and measured data, it is known that distant from the feeding point or close to the reactor walls poorly mixed zones with very low nutrient concentrations exist. An early scale-down study with *E. coli* employing oscillatory feeding protocols revealed the involvement of the stringent response in the cellular reaction to transient glucose starvation (Neubauer *et al.*, 1995a).

The stringent response is a global regulatory program usually preparing *E. coli* for entry into the stationary phase (Magnusson *et al.*, 2005; Gaca *et al.*, 2015; Hauryliuk *et al.*, 2015). Its hallmark is the synthesis of the alarmone (p)ppGpp on short time-scales by the ribosome-associated protein RelA or on longer time-scales by the bifunctional enzyme SpoT (Gallant *et al.*, 1970; Atherly, 1979; Murray and Bremer, 1996). ppGpp acts primarily as a transcription factor by binding to RNA polymerase and modulating its affinity to transcription initiation sites and alternative sigma factors. Additionally, ppGpp directly modulates the activity of certain proteins (Dalebroux and Swanson, 2012; Kanjee *et al.*, 2011).

The fast and reversible initiation of the stringent response to oscillatory substrate supply was later confirmed by measurements of ppGpp in continuous glucose chemostat cultivations in a two-compartment stirred tank-plug flow reactor (STR-PFR) setup (Löffler *et al.*, 2016). The feeding point was placed in the STR creating a starvation zone in the PFR, which allowed to resolve the timescale of cellular response. Moreover, it was shown that extensive transcriptional responses take place as cells move transiently through a nutrient poor zone. From theoretical calculations of ATP costs Löffler *et al.* estimated that an increase in maintenance energy demand of more than 30% was caused by the repeated exposure of cells to the nutrient gradient offering a new explanation for performance losses in large-scale bioprocesses (Löffler *et al.*, 2016). Analogous experiments with ammonium as the limiting nutrient revealed similar, yet less pronounced, regulation patterns affirming the importance of the stringent response for global regulation in *E. coli* in a scenario of oscillating starvation stimuli (Simen *et al.*, 2017). Fed-batch processes limited by ammonium or other nitrogen sources are interesting fermentation scenarios for the production of small molecules which mainly consist of carbon such as fatty alcohols (Chubukov *et al.*, 2017). Nitrogen limitation is commonly used to enhance the accumulation of cellular carbon storage products such as polyhydroxyalkanoates used for bioplastic synthesis (Wen *et al.*, 2010; Oliveira-Filho *et al.*, 2019), including *E. coli* as a potential host (Wang *et al.*, 2009). As nitrogen forms a relatively large part of cells, nitrogen limitation can be easily explored during process development. During scale-up, such processes will likely suffer from similar issues as carbon-limited processes (Simen *et al.*, 2017).

Recently, the strains *E. coli* SR and *E. coli* HGT with modulated stringent response were constructed in our laboratory (Michalowski *et al.*, 2017). The strains lack *relA* which is primarily responsible for rapid ppGpp synthesis upon nutrient depletion and carry modifications in the bifunctional enzyme SpoT. It was shown that they do not react to the exhaustion of ammonium supply by ppGpp synthesis (Michalowski *et al.*, 2017). Strain *E. coli* SR displays no negative phenotypic differences in batch cultivations compared to its parent strain *E. coli* K-12 MG1655. However, under conditions of ammonium limitation, *E. coli* SR was found to have an elevated specific glucose consumption rate which is beneficial for two-stage processes involving product formation in the nitrogen limited phase (Jarmander *et al.*, 2015; Perez-Zabaleta *et al.*, 2016).

The combination of properties displayed by *E. coli SR* indicates that this strain can potentially be developed as a platform strain for robust scale-up from lab to production. In this work, we compared the phenotypic and transcriptional responses of *E. coli* SR and its parent strain *E. coli* MG1655 in a two-compartment scale-down reactor. We focused our investigation on the regulatory differences between these strains in the response to repeated short-term stimuli. The primary stirred tank reactor was operated as an ammonium-limited chemostat while a plug flow reactor simulated a nitrogen starvation zone.

Results

Continuous cultivation with periodic nutrient depletion

We cultivated *E. coli* SR and *E. coli* MG1655 in two independent continuous fermentations each in a previously described scale-down reactor consisting of a primary stirred tank reactor (STR) and a secondary plug-flow reactor (PFR), schematically shown in Fig. 1 (Löffler *et al.*, 2016; Simen *et al.*, 2017; Ankenbauer *et al.*, 2020). *E. coli* SR is a strain with modulated stringent response that was engineered to alleviate the induction of the stringent response and the general stress response upon nutrient depletion (Michalowski *et al.*, 2017). The chemostat was operated at a dilution rate of $D = 0.2\ h^{-1}$ and ammonium was chosen as the limiting

nutrient. After establishment of a steady state in the STR alone, a reference sample (S0, t = 0 h) was taken and the PFR connected. Periodic passage from the STR (average residence time $-\tau_{STR} = 6.2min$) through the PFR (average residence time $-\tau_{PFR} = 2.6min$) then created a repeated short nitrogen starvation stimulus. The average residence times represent worst-case scenarios that are still consistent with mixing studies (Vrábel *et al.*, 2000; Noorman, 2011) and the volume ratio STR to PFR was approximately 3:1 to represent existing simulation results (Lapin *et al.*, 2006; Haringa *et al.*, 2017). The long-term response of cells was investigated from additional samples taken from the STR shortly after connection of the PFR (S5, t = 5 min) and after establishment of a new steady-state (S28, t = 28 h) in the two-compartment cultivation. The short-term response of cells to the PFR stimulus was monitored by sampling from five ports along the primary axis of the PFR at identical timepoints. Transcript samples for the PFR were taken from port 5 (P5_5 and P5_28).

Basic growth and fermentation data confirmed earlier results that there are no detrimental differences in fundamental physiological parameters (Table 1) between *E. coli* MG1655 and *E. coli* SR under nitrogen-limited conditions (Michalowski *et al.*, 2017). There were no statistically significant differences in any parameter (two-tailed t-test, $p > 0.1$). Both strains reached practically identical biomass yields on ammonium and depleted ammonium to equally low levels regardless of process time and PFR action (Fig. 2). The most noteworthy difference between *E. coli* MG1655 and *E. coli* SR was a reduced concentration of excess glucose in the fermentation broth of *E. coli* SR. Consequently, we calculated a lower biomass yield on glucose for *E. coli* SR (Table 1). Under conditions of long-term nitrogen starvation in batch fermentations *E. coli* SR had previously displayed a relaxation in glucose and nitrogen uptake coupling and we thus suspected an increased specific glucose uptake rate (Michalowski *et al.*, 2017). The calculated specific glucose uptake rate was higher for *E. coli* SR, but the difference was not statistically significant in our experiments (two-tailed *t*-test, *P*-value > 0.1). Data from the fermentation broth supernatant showed that both strains converted comparable amounts of substrate into acetate as the primary byproduct. Carbon balancing revealed an increased fraction of unknown substances among the fermentation products of *E. coli* SR which were identified as dissolved organic substances in the fermentation supernatant by total dissolved carbon analysis. The elevated glucose uptake rate of *E. coli* SR likely leads to higher byproduct formation of typical overflow metabolites such as lactate, pyruvate, formate and the regulator 2-oxoglutarate, all of which are known to accumulate under nitrogen-limited conditions with glucose excess (Hua *et al.*, 2004). Apart from the primary byproduct acetate, individual small carbon byproducts were not measured as the overall total carbon efflux/influx balancing was in good agreement for both strains. Carbon recovery was 101 ± 2 % for *E. coli* MG1655 and 102 ± 1 % for *E. coli* SR indicating that in sum all relevant substances were detected.

In general, process time and the periodic PFR stimulus hardly affected global process parameters which is in accordance with former observations made in this reactor setup for nitrogen limitation and K-12 strains (Simen *et al.*, 2017). In sharp contrast, we found substantial regulatory differences between the two strains both in the short-term and in the long-term transcriptional responses to the periodic starvation stimulus.

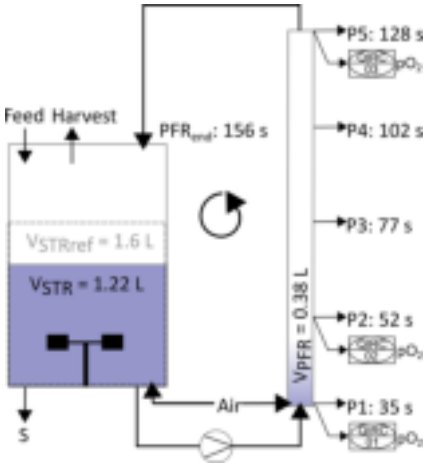


Fig. 1. Experimental design of the two-compartment system. The fermenter consists of a stirred tank reactor (STR) as the primary cultivation vessel and a plug-flow reactor (PFR) connected by an active pump. The ammonium-limited chemostat was operated at a dilution rate of D = 0.2 h^{-1} with the feeding point placed in the STR. The STR served as a limitation zone and the PFR formed a starvation zone. The setup was designed to resolve different timescales of cellular response. Oxygen saturation was measured by three oxygen probes and recorded by the process control system (01, 02, 03). V_{STRref}: Reference Volume without connection of PFR (constant volume).

Transcriptomic analysis: Overview

RNA-seq-based transcriptomic data to examine potentially important genes for the ammonium stress response of *E. coli* WT and *E. coli* SR was analysed. After filtering, 4037 predicted *E. coli* genes remained for further

Table 1. Physiological measurements.

E. coli MG1655	E. coli SR	
$Y_{XN}\left[\frac{g_{CDW}}{g_{NH_4^+}}\right]$	4.63 ± 0.12[a]	4.62 ± 0.27
$Y_{XS}\left[\frac{g_{CDW}}{g_{Glucose}}\right]$	0.32 ± 0.01	0.28 ± 0.01
$c_{Glucose,STR}\left[\frac{g_{Glucose}}{l}\right]$	2.07 ± 0.25	1.49 ± 0.06
$c_{Acetate,STR}\left[\frac{g_{Acetate}}{l}\right]$	1.39 ± 0.11	1.29 ± 0.14
$q_{NH_4^+}\left[\frac{g_{NH_4^+}}{g_{CDW}*h}\right]$	0.04 ± 0.01	0.05 ± 0.01
$q_S\left[\frac{g_{Glucose}}{g_{CDW}*h}\right]$	0.63 ± 0.05	0.77 ± 0.14
$q_{Ac}\left[\frac{g_{Acetate}}{g_{CDW}*h}\right]$	0.10 ± 0.01	0.10 ± 0.01
$q_{CO_2}\left[\frac{mmol_{CO_2}}{g_{CDW}*h}\right]$	8.73 ± 1.06	9.98 ± 2.23
$q_{O_2}\left[\frac{mmol_{O_2}}{g_{CDW}*h}\right]$	9.28 ± 0.47	10.9 ± 2.02
$RQ\left[\frac{mol_{CO_2}}{mol_{O_2}}\right]$	0.95 ± 0.16	0.91 ± 0.04
$q_{ATP}\left[\frac{mmol_{ATP}}{g_{CDW}*h}\right]$	29.23 ± 0.62[b]	34.73 ± 6.39
$D\left[\frac{1}{h}\right]$	0.20 ± 0.01	0.21 ± 0.03

a. Errors indicate SEM ($n = 2$). All rates were calculated from averaged values collected over the entire STR-PFR process time.
b. Estimated values assuming a P/O-Ratio of 1.2.

analysis (see Supporting information). The fast tactical transcriptional response to ammonia shortage was determined by comparing PFR port 5 samples to STR samples taken at the same process time points. Long-term responses were studied by comparing post-perturbation samples from the STR after 5 min (S5) and 28 h (S28) to the reference sample (S0). The statistical threshold for significance was set for adjusted p-value < 0.01 and log2FC > |1|. 54 differentially expressed genes (DEGs) (UP: 14, DOWN: 40) formed the long-term response of *E. coli* MG1655. The short-term response was more pronounced comprising 837 DEGs (UP: 242, DOWN: 595). *E. coli* SR disclosed a similar number of 61 DEGs for the long term response (UP: 12, DOWN: 49), but substantially less DEGS as short term response (Total: 387, UP: 161, DOWN: 226) (Fig. 3). Log_2FC values range from −4.69 to 4.96 (WT) and −3.90 to 5.13 (SR). Fig. 3 depicts an overview of transcriptional dynamics outlining the halved response of *E. coli* SR 5 min after repeated nitrogen limited perturbation compared to WT.

Figure 4 shows that the multi-transcript response of each strain could be well described by 2-dimensional PCA covering 96% and 87% of total variance for *E. coli* WT and *E. coli* SR, respectively. Notably, biological duplicates were found in close proximity. PC1 accounts for the sample port location, PC2 for the time course. Unique and clearly distinguishable differences between STR and PFR transcript patterns were observed already after 5 min of repeated nitrogen starvation for both strains (Figure 4, A1). In particular, principal component 1 (PC1) disclosed major differences between the samples of each strain accounting for 88% and 67% regarding *E. coli* WT and *E. coli* SR, respectively. The PCA finding is in agreement with the reduced number of DEGs observed for *E. coli* SR. The impact of PC2 is more pronounced for *E. coli* SR although almost

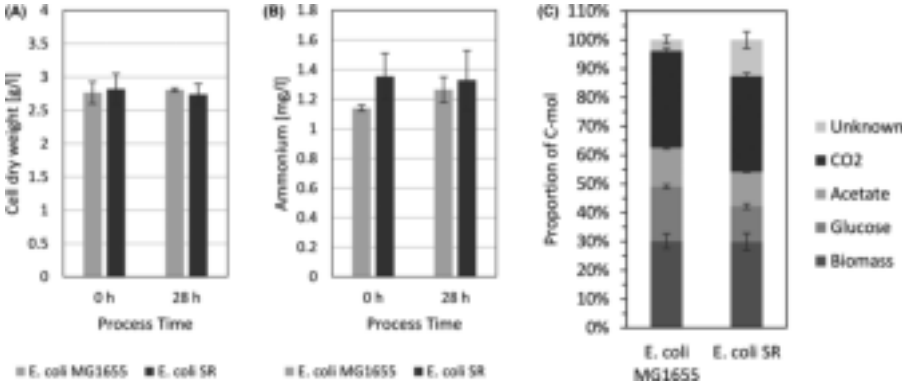


Fig. 2. Physiological measurements.A. Cell dry weight. Concentration of cell dry weight after at least 25 h chemostat process before connecting the plug-flow reactor (0 h) and after 28 h of chemostat process with connected PFR (28 h).B. Ammonium. Concentration of residual ammonium in the supernatant.C. Carbon Balance. Columns show efflux fractions of total C-mol based on carbon influx. The final fraction represents undetermined dissolved organic substances in the fermentation broth, as measured by the difference of all efflux carbon detected by exhaust gas or total carbon analysis and the sum of the individually measured efflux components. Error bars indicate SEM ($n = 2$) of individual components (A, B and C).

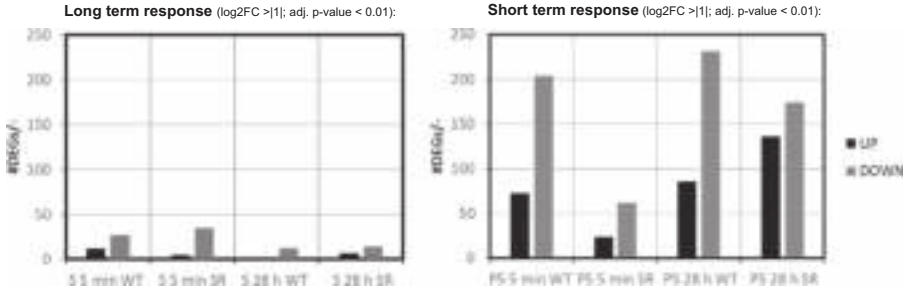


Fig. 3. Number of UP (black) and DOWN (gray) regulated genes (DEGs). Long-term (left) and short-term (right) response to repeated nitrogen starvation for *E. coli* MG1655 (WT) and *E. coli* SR (SR) and given process times.

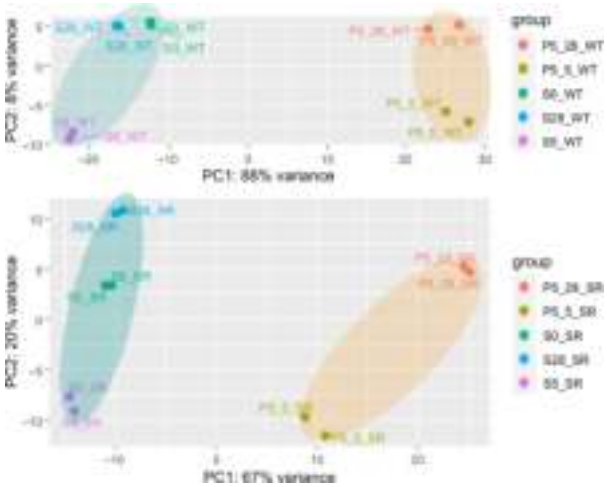


Fig. 4. Principal component analysis of transcript data of *E. coli* MG1655 (WT) (top) and *E. coli* SR (bottom) obtained from STR (S) and PFR (port 5, P5) at three process time points (0 h, 5 min, and 28 h). Covered measurement variance of each principal component (PC) is indicated. Ellipses cluster samples of STR and PFR. PC1 accounts for 'sample port location', PC2 for 'process time'.

identical numbers of DEGs were found as long-term response in both strains. However, given the low impact of PC1 for *E. coli* SR, similar DEG values affect the relative principal component analysis stronger.

As long-term responses of both strains were similar (see Appendix: Supporting information) and weaker than short-term responses (Fig. 4) further analysis focused on short-term transcript patterns. Notably, changes between long- and short-term responses of both strains were dominated by counteracting transcript dynamics resetting perturbations after PFR passages (MG1655: 5 min and 28 h). Observations are in line with similar findings (Chang *et al*., 2002). Additional differences were found in the upregulation of carbohydrate transport (SR: 5 min) and catabolic processes (SR: 28 h) (see Fig. A5 and A6).

Regulatory response to short-term ammonium limitation

Preceding investigations of *E. coli* K-12 strains in STR-PFR scale-down reactors revealed the rapid accumulation of the alarmone ppGpp upon entry into the nutrient

limited zone under both glucose and ammonium limitation (Löffler *et al.*, 2016; Simen *et al.*, 2017). Concomitantly, an extensive transcriptional reprogramming of cells occurred. In standard batch fermentations *E. coli* SR in turn did not react to ammonium depletion by ppGpp synthesis (Michalowski *et al.*, 2017). We therefore measured intracellular ppGpp levels from samples taken from the five ports of the PFR along its primary axis (Fig. 5). During the PFR passage *E. coli* MG1655 accumulated ppGpp to levels 2 – 3 fold higher than measured in the STR, displaying the same behaviour as previously observed for the closely related K-12 strain *E. coli* W3110 (Simen *et al.*, 2017). In contrast, *E. coli* SR had no elevated levels of ppGpp at any point during the PFR passage regardless of process time. These results complement previous findings for the case of repeated short stimuli and confirm the strain's resilience to ammonium exhaustion.

Based on these encouraging findings, we focused our investigation on the short-term transcriptional response of both strains along the PFR axis. We compared data from samples drawn from port 5 of the PFR to samples drawn from the STR at identical process time points. Short-term changes revealed a significantly different response of *E. coli* SR compared to *E. coli* MG1655 not only in the amount of DEGs (Fig. 3), but also in the function of these genes (Fig. 4, 6). To elucidate patterns in the transcriptional responses, we searched for common DEGs, investigated the behaviour of gene clusters of orthologous groups (COGs), and compared sigma factor (σ) activities. The gene expression patterns of each strain individually were assigned to 21 functional categories based on the COG database (Tatusov *et al.*, 2003). In total 3532 of the 4037 genes (87.5%) could be annotated to COG. For each COG category, the resulting t-values are represented in a lollipop plot (Fig. 7). Significant changes were defined with a FDR-corrected p-value < 0.01. Furthermore, the activation and deactivation of sigma factors over time were investigated (Fig. 7). In this case, 3935 out of 4037 genes could be assigned to the sigma factor-gene interaction database from RegulonDB (Santos-Zavaleta *et al.*, 2019).

After the first 5 min of PFR action *E. coli* MG1655 and *E. coli* SR exhibited substantially different transcriptional responses. The strains had only 64 DEGs in common, split equally between up- and down regulation (Fig. 6 left). Hence, these genes mirror the transcriptional response to short-term starvation irrespective of a functional stringent response, in which 14 out of the 32 common upregulated genes are associated with the Ntr-reponse (e.g. *glnK*, *amtB*, *glnAHPQ*, *rutA*). Downregulated genes consist of genes responsible for amino acid biosynthesis (e.g. *argCF*, *metABFINR*) and other cellular functions such as DNA cleavage, transporters and oxidoreductases. The only oppositely regulated gene was *guaC* coding for the GMP reductase GuaC. Transcriptional control of the *guaC* promoter by the stringent response was proposed after its initial discovery and is clearly supported by our data (Andrews and Guest, 1988). Individual, strain-specific short-term regulation was observed for 398 (*E. coli* MG1655) and 32 (*E. coli* SR) specific DEGs after 5 min, clearly demonstrating the effect of the stringent response on the *E. coli* transcriptome.

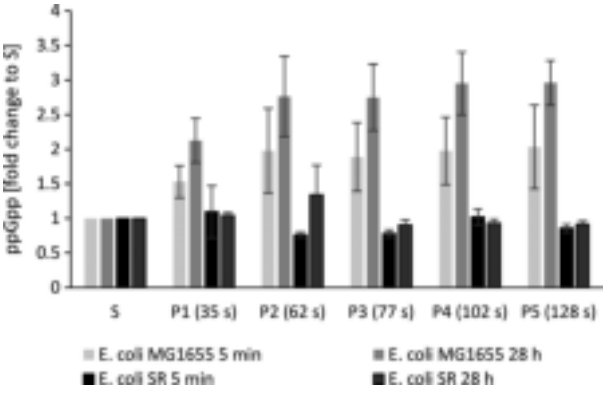


Fig. 5. Alarmone accumulation along the PFR. Concentration of ppGpp measured from samples drawn along the plug flow reactor (P1 to P5) relative to the concentration measured in the stirred tank reactor (S, all values set to 1) for *E. coli* MG1655 (WT) and *E. coli* SR (SR). Error Bars represent SEM ($n = 2$).

Fig. 6. Venn diagrams representing partially overlapping sets of DEGs of *E. coli* MG1655 (WT) and *E. coli* SR. The number of significantly up- (UP) and downregulated (DOWN) genes in each set is indicated by numbers. Left: Short-term responses 5 min after PFR connection. Right: Short-term responses 28 h after PFR connection. Complete gene lists of the Venn diagrams are available in the supplementary data.

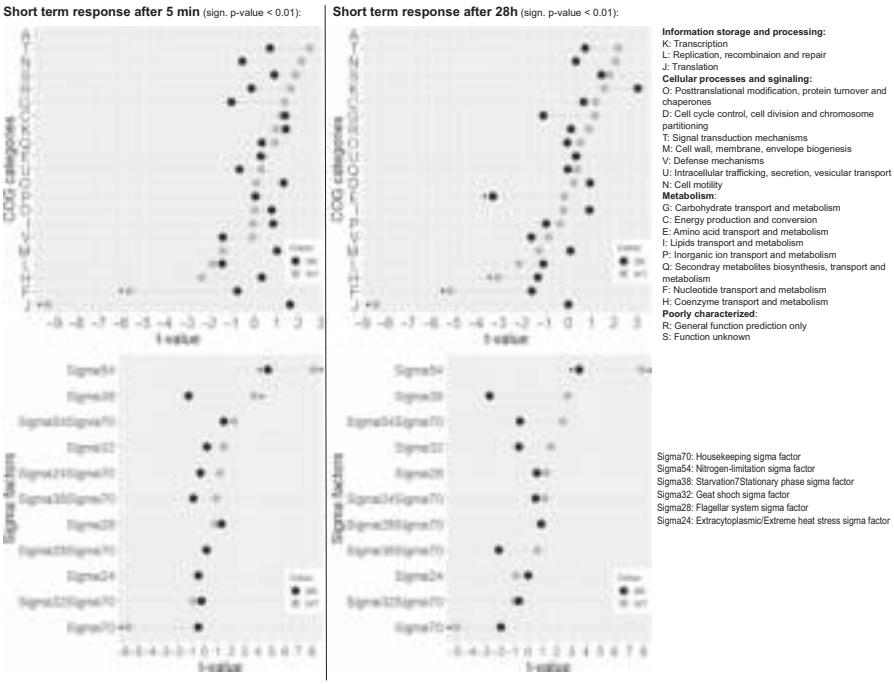


Fig. 7. Top: Transcriptional patterns grouped into COG categories of *E. coli* MG1655 (WT) and *E. coli* SR (SR). Left: short-term patterns to the PFR stimulus 5 min after PFR connection. Right: short-term patterns to the PFR stimulus 28 h after PFR connection. Bottom: Sigma factor activities of *E. coli* MG1655 (WT, grey) and *E. coli* SR (SR, black). Left: Short-term response to the PFR stimulus 5 min after PFR connection. Right: Short-term response to the PFR stimulus 28 h after PFR connection. Significant categories are indicated with an asterix.

Gene expression along the PFR after 28 h of PFR action differs strongly from the early response. 125 DEGs, mostly downregulated, are shared by both strains and the number of individually regulated genes is similar with 242 genes for *E. coli* MG1655 and 158 genes for *E. coli* SR (Fig. 6 right). Additionally, seven genes are oppositely regulated. Three of them (*tolQ, guaC, purM*) are upregulated in *E. coli* SR and downregulated in

E. coli MG1655. These genes correspond to cell envelope integrity during cell division (Gerding *et al.*, 2007), nucleotide metabolism (Kanjee *et al.*, 2012) and purine *de novo* biosynthesis (Mueller *et al.*, 1999). While purine *de novo* biosynthesis is actively inhibited by ppGpp via inhibition of GuaB, GTP synthesis solely originates from purine salvage pathways with *xdhA* significantly increased in *E. coli* MG1655 (Xi *et al.*, 2000). The residual four oppositely regulated DEGs (*csiD, glnL, lhgO, yeaH*) predominantly play a role in the adaptation to nitrogen starvation and except for *glnL* are known to be induced by ppGpp. NtrB encoded by *glnL* is an essential part of the Ntr response cascade to nitrogen starvation and *yeaG* positively impacts *rpoS* transcription and translation under prolonged nitrogen starvation (Brown *et al.*, 2014). Despite these differences in adaption to nitrogen limitation, we observed no alterations in the uptake or utilization of ammonium which indicates that the additional regulatory adaptions of *E. coli* MG1655 are irrelevant in the context of a bioprocess.

Transcriptional patterns could be identified by functional enrichments of groups based on COG categories and sigma factor activities. COG groups J (Translation, ribosomal structure, and biogenesis) and F (nucleotide transport and metabolism) were significantly down regulated as part of the stringent response of *E. coli* MG1655 after both 5 min and 28 h (Fig. 7). For the 28 h sampling point group H (coenzyme transport and metabolism) was also significantly downregulated. As already indicated by the oppositely regulated genes (Fig. 7), σ54-mediated genes responsible for the activation of the Ntr stress response including *yeaG/H* via NtrBC were induced in *E. coli* MG1655, as well as the σ38 regulon as part of the general stress response (Brown *et al.*, 2014; Figueira *et al.*, 2015) (Fig. 7). Due to the limited amount of RNA-Polymerase (RNAP) core enzymes, σ70 competes with σ54, resulting in an antiproportional expression of their mediated genes (Jishage *et al.*, 1996). In contrast, *E. coli* SR only increased the expression of genes regulated by σ54 after 5 min and no significant COG category was identified at this time-point. The absence of the stringent response in *E. coli* SR is clearly visible in an overall dampened regulatory response. The only significantly regulated group is E (amino acid transport and metabolism) after 28 h of PFR action, and the significantly downregulated genes in this group are predominantly ABC-transporters.

To unravel more detailed patterns in the transcriptional responses we assigned genes to the up-to-date gene ontology (GO) gene sets using GAGE (Luo *et al.*, 2009). 3345 out of 4037 genes (83%) could be mapped to GO Terms. As shown in Fig. 3 the majority of significant DEGs for *E. coli* MG1655 were downregulated. This is mirrored by the results of the identified top 20 GO categories which were uniformly down-regulated (Fig. 8). *E. coli* MG1655 predominantly downregulated genes related to ribosomal biosynthesis and translation after 5 min and 28 h as expected for a stringent phenotype (Fig. 8). These transcriptional changes are counteracted in the long-term response observed from the STR (Fig. A3 to A6) which indicates looping induction and repression of the genes. Patterns from *E. coli* SR were less pronounced and grouped differently. After 5 min we observed decreasing gene expression of ATP-demanding processes such as ABC transporters and ATPase complexes (Fig 8). After 28 h the PFR passage mainly induced an increased negative regulation of transcription and metabolic processes (Fig. 8). Care must be taken in the interpretation of this group though. General categories affecting transcription (GO:0006351, GO:0045892, GO:0097659, GO:1903507) or RNA processes (GO:0032774, GO:1902679, GO:0051253, GO:0051252) are represented as simultaneously negatively and positively regulated. Moreover, all negative regulators included in these terms, such as members of the CRP family, are also capable of positive regulation. Other negative regulation categories involve genes which actively inhibit translation and belong to SOS signals like DNA damage, prevention of cell division and programmed cell death (PCD). *E. coli* SR thereby focuses on σ38 regulated genes, as well as toxin and antitoxin systems (*mazEF* and *mqsRA*) possibly resulting in arrested growth and a dormant cell state or even PCD. As growth arrest is usually a primary outcome of the stringent response, which is absent in *E. coli* SR, we hypothesize that this pattern might provide an alternative way for *E. coli* SR to achieve cell cycle arrest.

In summary, the short-term response transcriptional patterns of *E. coli* MG1655 were extensive and dominated by the stringent response and the Ntr regulon. The major activated sigma factors were σ54 and σ38. Overall, the transcription of ribosomal genes and other genes necessary for growth was inhibited, while genes involved in the transport and fixation of ammonia were induced. Our observations reflect well-known regulatory patterns exerted by *E. coli* K-12 when facing nitrogen starvation (Chang *et al.*, 2002; Traxler *et al.*, 2008; Traxler *et al.*, 2011; Simen *et al.*, 2017; Wang and Levin, 2009). In contrast, the transcriptional short-term response of *E. coli* SR is dampened both in the number of DEGs and the patterns observed, especially shortly after connection of the PFR. The only significantly activated sigma factor is σ54 indicating a functional but attenuated Ntr response in the absence of ppGpp accumulation. Adaptation to ongoing starvation was possibly attempted via negative regulation of metabolic processes and SOS pathways.

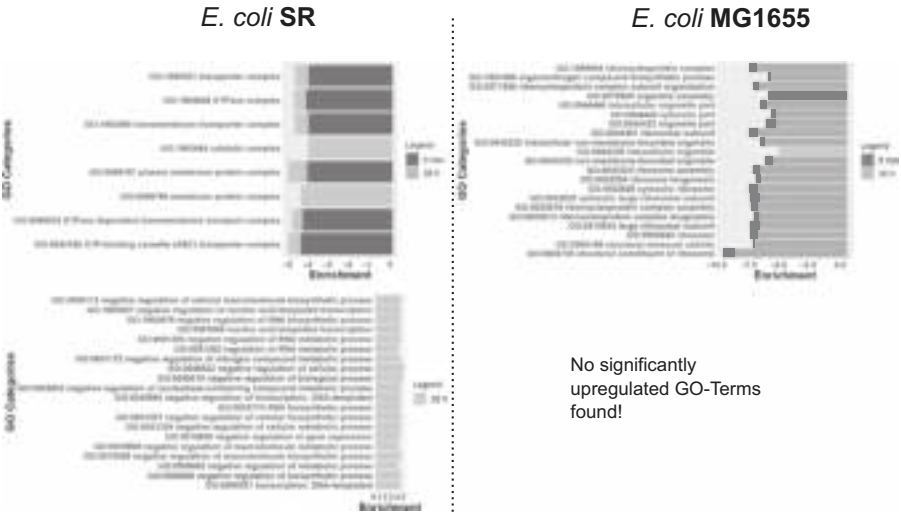


Fig. 8. Significant GO categories after 5 min and 28 h of both *E. coli* SR (left) and *E. coli* MG1655 (right). Downregulated categories are arranged at the top and upregulated GO terms at the bottom. 5 min: Short-term response of *E. coli* SR (left) and *E. coli* MG1655 (right) after 5 min of PFR action. Only the Top 20 out of 102 significantly downregulated categories are shown. Neither strain had significantly upregulated categories for this time-point. 28h : Short-term response of *E. coli* SR (left, light grey) and *E. coli* MG1655 (right, light grey) after 28 h of PFR action. For *E. coli* SR only the Top 20 out of 24 significantly upregulated categories are shown. For *E. coli* MG1655 only the Top 20 out of 95 significantly downregulated categories are shown. No significantly upregulated categories were found for this time-point.

Discussion

In the present study, we investigated the regulatory responses of the stringent response mutant strain *E. coli* SR when exposed to repeated short starvation stimuli in a scale-down reactor. The comparison with its wild-type parent *E. coli* MG1655 unravelled dampened regulatory patterns which are potentially beneficial for the application of *E. coli* SR in industrial large-scale reactors. The reduced regulatory patterns might be beneficial for heterologous protein expression as well as the production of small molecules as less interference with engineered metabolic pathways may occur and energy otherwise spent for adaptive responses is available for product formation.

An important finding of our study is that despite the regulatory differences *E. coli* SR displayed no dysfunctionalities in handling the shortage of ammonium. *E. coli* SR reached the same biomass yield on ammonium as *E. coli* MG1655 both with and without PFR action. Moreover, both strains depleted ammonium to comparable levels of about 1.2 mg l^{-1} or 67 μmol l^{-1}, well in line with previously reported values for *E. coli* K-12 strains in nitrogen limited chemostats (Hua *et al.*, 2004). The low remaining ammonium concentration indicates that uptake in both strains is mediated actively by AmtB with K_m = 0.8 mM (Williamson et al., 2020) and incorporation is accomplished by the GS-GOGAT System with GS K_m = 0.1 mM (Alibhai and Villafranca, 1994). This is supported by our transcriptional data which revealed that *amtB*, *gltB* and *gltD* were significantly enriched for both strains over all time-points. Transcripts of *glnA* were also always significantly enriched except for the time point 28 h of *E. coli* SR. Concomitantly, we identified transcriptional patterns typical for the σ54- and NtrBC-mediated responses to nitrogen starvation (Reitzer, 2003). 13 out of 21 known NtrC-regulated operons (Brown *et al.*, 2014) were induced at PFR port 5 in *E. coli* MG1655 at all time points (Table S2). For *E. coli* SR, the Ntr response was slightly reduced, with 9 out of 21 operons induced (Table S3) and lower overexpression of σ54 transcribed genes. These findings lead to the conclusion of an active, but diminished Ntr response of *E. coli* SR that still allowed fully functional ammonium assimilation. Additionally, the energy consumption as maintenance add-on for both strains was calculated according to Löffler *et al.* (2016) assuming *de novo* synthesis of all upregulated DEGs over the whole process time (28 h).

The resulting energy savings of *E. coli* SR due to weaker transcriptional response added up to around 46.5 %. In terms of microbial productivity, the reduced maintenance demand potentially increases the amount of available ATP for biomass-specific productivities and improves cell fitness.

In a previous study, a significantly elevated specific glucose consumption rate under ammonium limitation was observed in *E. coli* SR (Michalowski *et al.*, 2017). Similarly, we observed reduced excess glucose and the accompanying formation of dissolved byproducts in the fermentation supernatant. In *E. coli* K-12 strains, the consumption of glucose is usually tightly coupled to the availability of nitrogen on the level of metabolite control by the interaction of 2-oxoglutarate with PtsI (Doucette *et al.*, 2011). The exact mechanism by which coupling of nitrogen and glucose uptake rates are relaxed in *E. coli* SR is not clear as the strain is isogenic to *E. coli* MG1655 except for the deletion of *relA* and the modifications in *spoT*. However, we found an increased transcription of *ptsI*, *ptsH* and *ptsG* in *E. coli* SR compared to *E. coli* MG1655 (Table S6). Artificially increased expression of *ptsI* has been shown to increase specific glucose uptake rates in nitrogen limited conditions (Chubukov *et al.*, 2017). We presume that the increased glucose uptake rate in *E. coli* SR might be caused by deregulated expression of *ptsI*, potentially connected to the absence of the stringent response by the action of CRP whose transcription is negatively regulated by ppGpp (Johansson *et al.*, 2000). It remains to be clarified whether *E. coli* SR has altered cytoplasmic 2-oxoglutarate levels or the action of ppGpp influences the coupling of glucose consumption to nitrogen availability, potentially by the proposed mechanism. Increased specific glucose uptake rates in conjunction with higher respiratory activity have also been observed in *E. coli* MG1655 subjected to repeated glucose feast-famine cycles (Vasilakou *et al.*, 2020). Future studies should thus examine how *E. coli* SR reacts to varying availability of glucose or other carbon sources.

In view of these differences in carbon metabolism, we hypothesized that biological energy availability might be unequal for *E. coli* MG1655 and *E. coli* SR. From oxygen and glucose uptake rates the specific ATP production rate q_{ATP} was estimated (Table 1). q_{ATP} greatly depends on the effective P/O ratio and current scientific consensus estimates realistic P/O ratios between 1.0 and 1.5 for *E. coli* (Noguchi *et al.*, 2004; Szenk *et al.*, 2017). For our estimations of q_{ATP} we assumed a conservative P/O ratio of 1.2 and 2 moles of ATP per mol glucose from glycolysis. The result indicates that *E. coli* SR might have an increased availability of ATP compared to its wild-type parent under the applied experimental conditions. Given that the respiratory capability and thus the ATP production capability of K-12 strains is not exhausted at a dilution rate of $D = 0.2\ h^{-1}$ it appears that the increased glycolytic flux to byproducts displayed by *E. coli* SR was also not a result of increased energy demand. Moreover, increased glucose uptake has been reported previously for *E. coli* SR under conditions of ammonia limitation despite high adenylate energy charge (Michalowski *et al.*, 2017). Carbon and redox homeostasis at elevated glycolytic flux would then be maintained by byproduct excretion and increased respiration, possibly involving the dissipation of surplus energy by uncoupling of the electron transport chain (Bekker *et al.*, 2009).

Nitrogen limitation inducing the stringent response is a well-documented phenomenon in *E. coli.* Multiple previous studies predominantly observed heavily increased gene expression corresponding to amino acid transport and metabolism (Barker *et al.*, 2001; Durfee *et al.*, 2008; Traxler *et al.*, 2008; Traxler *et al.*, 2011; Brown *et al.*, 2014; Simen *et al.*, 2017). Conversely, we observed almost equally distributed up- and downregulated genes for amino acid transport and metabolism (see Supporting information: Transcriptomics), which was only reported by few research groups (Chang *et al.*, 2002; Traxler *et al.*, 2008). As a result, no overall significant statistical trend was detectable for this category (Fig. 7). We suggest that the individual operons do not solely respond to ppGpp, but rather depend on other signals and regulatory networks which were not found to be significantly expressed in this study such as the Lrp regulon. Additionally, caution is advised when comparing transcriptomic analyses originating from different studies as they greatly depend on the transcriptional reference state and thus the details of the experimental design.

In general, the amount of DEGs of *E. coli* K-12 MG1655 was similar to the numbers found in the analogous study of Simen *et al.* (2017) who employed the closely related *E. coli* K-12 W3110 confirming the validity of our data. The amount of DEGs is also less than observed during the related study of glucose starvation by Löffler *et al.* (2016) which points towards significant potential of *E. coli* SR to preserve energy in glucose starvation conditions. An interesting difference to the former studies in this scale-down reactor setup is the absence of increased motility in the STR after PFR connection (Löffler *et al.*, 2016; Simen *et al.*, 2017). Our dataset contains no upregulated flagellar or sigma factor 28 mediated gene patterns from the STR at any timepoint (Fig. 7). We first hypothesized that the cause might be genetic differences affecting motility which are well documented between MG1655 and W3110 and even between different MG1655 isolates (Barker *et al.*, 2004; Hayashi *et al.*, 2006). However, sequencing of our MG1655 isolate revealed the presence of the canonical

IS-1 insertion upstream of *flhD* which confers motility and our MG1655 isolate displayed vivid spreading in motility agar (Supporting information, Fig. A7). An alternative explanation could be derived by the interplay of quorum sensing and flagellar regulation through the action of autoinducer-2 (AI-2) and the motility quorum sensing regulator MqsR. While transcript levels of *luxS* (LuxS synthesizes AI-2) remain unchanged, the expression of *mqsR* is significantly enriched at PFR port 5 and MqsR is known to induce the flagellar synthesis cascade (González Barrios *et al*., 2006). However, cell dry weight (CDW) was always below 3 g l^{-1} in our experiments whereas Simen *et al.* worked with around 10 g l^{-1} CDW. Higher biomass should lead to increased AI-2 levels and may cause a preconditioned phenotype that rapidly initiates flagellar biosynthesis when encountering nutrient stress. Thus, rapid induction of motility genes might become more pronounced during high cell density processes in large-scale reactors and remains to be examined in further studies. Additionally, as introduced by Löffler *et al.*(2016) during glucose fluctuation, genes of the category cell motility were identified as one of the most prominent energy consumers and might therefore be candidates for genome reduction (Löffler *et al*., 2016).

Analysis of gene expression patterns (Fig. 7 and 8) revealed that both strains individually adapted to repeated nitrogen starvation. *E. coli* MG1655 adjusted by utilizing the ppGpp-mediated general stress response including activation of toxin/antitoxin (TA) systems like *mqsRA* and *mazEF*. This strategy intends to arrest the cell cycle and form persister cells (Balaban *et al*., 2004). Persister cell formation is not yet fully understood and usually only involves a small fraction of cells (Chowdhury *et al*., 2016; Gerdes and Maisonneuve, 2012; Korch *et al*., 2015). Thus, it seems to be only of minor importance for industrial processes but some persister genes affect persister level due to altered growth rates rather than contributing to a mechanism of cell cycle arrest and might have a significant impact on bioprocess performance (Allison *et al*., 2011). Nonetheless two common dependencies affecting persister formation, ppGpp and TA systems, are known which is in line with our findings (Aizenman *et al*., 1996; González Barrios *et al*., 2006; Chowdhury *et al*., 2016; Sun *et al*., 2017; Wang and Levin, 2009). Persister formation benefits from increased ppGpp concentrations but is still possible at lower rates in the absence of ppGpp by proteins which simply reduce growth (Chowdhury *et al*., 2016). The nucleotide pyrophosphohydrolase MazG which is negatively regulated by the *mazEF* system is able to initiate cell cycle arrest and was significantly upregulated in *E. coli* SR after 28 h (Lee *et al*., 2008). Additionally, *E. coli* SR initiated negative regulation of transcription, translation and cell division processes as part of the SOS response (Fig. 8). Most likely, the SOS pathways were activated due to ongoing DNA replication during starvation conditions which might ultimately result in DNA damage and inhibited cell division (Bi and Lutkenhaus, 1993; Joseleau-Petit *et al*., 1999; Traxler *et al*., 2008). As part of the SOS response and as a key gene involved in filamentation *sulA* was significantly upregulated in *E. coli* SR. SulA inhibits the initiation of cellular division by repressing the assembly of FtsZ into the Z ring (Huisman *et al*., 1984; Fonville *et al*., 2010). Simultaneously with the overexpression of *sulA*, *lexA* was significantly increased which acts as a major repressor of SOS signals. LexA regulates the response strength and is actively involved in the occurrence of persister cells in bacterial populations (Butala *et al*., 2011). These results indicate a coordinated and rather complex SOS response in *E. coli* SR to form persister cells which is not yet fully understood.

The natural regulation of *E. coli* has evolved towards optimality in its lifestyle as a gut bacterium and is not honed for the demands of a large-scale bioprocess. The absence of the stringent response and the conservation of the ability to grow efficiently in minimal medium suggest that *E. coli* SR has the potential to become a platform strain for applications in large-scale reactors. Our transcriptional analysis shows that the short-term response of *E. coli* SR to ammonium depletion is dampened but a functional Ntr/σ54 response remains. Regarding glucose-limited fermentations, we hypothesize that *E. coli* SR has significant potential to preserve energy in such conditions since the regulatory responses are usually even more pronounced and centred around the stringent response (Hardiman *et al*., 2007; Löffler *et al*., 2016). We therefore propose to confirm the suitability of *E. coli* SR for large-scale applications in multi-compartment scale-down reactors employing exemplary small-molecule production scenarios. These should include standard glucose-limited fed-batches as well as ammonium limited fed-batches with a prolonged nitrogen-limited production phase to exploit its elevated glucose consumption.

Experimental procedures

Bacterial strains and media

Strains *E. coli* MG1655 or *E. coli* SR were used in all experiments (Table 2).

2xYT agar plates were prepared by autoclaving 16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl and 18 g l^{-1} agar-agar dissolved in demineralized water. Minimal medium for precultures consisted of 4 g l^{-1} glucose, 0.96 g l^{-1} $NaH_2PO_4 \cdot 2H_2O$, 3.51 g l^{-1} K_2HPO_4, 2.4 g l^{-1} $(NH_4)_2SO_4$, 0.01 g l^{-1} thiamine hydrochloride and 0.2%

Table 2. Bacterial Strains used in this study.

Strain	Genotype/strain information	Reference
Escherichia coli K-12 MG1655 ("wild type" strain, abbrev. WT)	F^-, λ^-, *ilvG*$^-$, *rfb*-50, *rph*-1	Michalowski et al. (2017)
Escherichia coli SR	MG1655 Δ*relA*, *spoT* [R290E;K292D]	Michalowski et al. (2017)

(V/V) trace elements stock solution. Minimal medium for batch cultivation in the bioreactor consisted of 19 g l^{-1} glucose, 1.50 g l^{-1} $NaH_2PO_4{\cdot}2H_2O$, 3.9 g l^{-1} K_2HPO_4, 5.7 g l^{-1} $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. 200 µl of antifoaming agent Struktol J647 (Schill + Seilacher, Hamburg, Germany) was added to the batch medium prior to inoculation. Minimal medium for continuous chemostat cultivation in the bioreactor consisted of 11.4 g l^{-1} glucose, 1 g l^{-1} $NaH_2PO_4{\cdot}2H_2O$, 2.6 g l^{-1} K_2HPO_4, 2.28 g l^{-1} $(NH_4)_2SO_4$ and 0.2% (V/V) trace elements stock solution. Throughout the chemostat phase 50 µl/h of antifoaming agent Struktol J647 were added continuously to the fermentation medium. The composition of trace element stock solution was 4.175 $FeCl_3{\cdot}6H_2O$, 0.045 g l^{-1} $ZnSO_4{\cdot}7H_2O$, 0.025 g l^{-1} $MnSO_4{\cdot}H_2O$, 0.4 g l^{-1} $CuSO_4{\cdot}5H_2O$, 0.045 $CoCl_2{\cdot}6H_2O$, 2.2 g l^{-1} $CaCl_2{\cdot}2H_2O$, 50 g l^{-1} $MgSO_4{\cdot}7H_2O$ and 55 g l^{-1} sodium citrate dihydrate. Stock solutions of salts, trace elements and glucose were autoclaved separately, and stock solutions of thiamine hydrochloride were filter sterilized and stored at 4°C. All compounds were combined just before the experiments to prevent possible aging of media.

Bioreactor setup

Cultivations were carried out in a two-compartment scale-down reactor. The primary reactor was a stirred tank reactor (STR), and a plug flow reactor (PFR) was used as the secondary compartment mimicking a starvation zone. The plug flow reactor was connected to the stirred tank reactor after establishment and sampling of a steady state in the chemostat phase. The basic technical setup has been characterized previously (Löffler *et al.*, 2016; Simen *et al.*, 2017). Minor modifications to the original setup have been made and are described elsewhere (Ankenbauer *et al.*, 2020).

The primary reactor was a 3 l bioreactor (Bioengineering, Wald, Switzerland) equipped with flow baffles and two six-blade Rushton type impellers operated at 1000 rpm. A constant aeration rate of 2.0 standard litres of ambient pressurized air per minute was employed and the system operated at a total pressure of 1.5 bar. Temperature was monitored by a platinum resistance thermometer and regulated by electrical heating or water cooling. Temperature was set to 28–30°C for the batch phase and to 37°C for the continuous chemostat phase. The reactor was equipped with a pH sensor (Mettler Toledo, Columbus, USA) to control pH and a pO2 sensor for monitoring dissolved oxygen tension (PreSens, Regensburg, Germany). During all fermentation stages pH was set to 7.0 and regulated by automated addition of 3 M NaOH or 2.5 M H_3PO_4. Dissolved oxygen tension was not regulated but maintained values above 70% saturation to 1.5 bar ambient air throughout the entire cultivation. In the exhaust gas stream, the concentration of oxygen and carbon dioxide was measured by gas sensors (BlueSens, Herten, Germany). During the chemostat phase the feed was constantly added to the reactor by a peristaltic pump (Watson-Marlow, Falmouth, UK). The feed flow was monitored by a balance recording the weight of the stirred feed barrel and manually adjusted if necessary. The harvesting pump operated as a slave pump set to maintain a constant weight of the bioreactor. For this purpose, the stirred tank reactor was installed on a balance as well.

The secondary compartment was a plug-flow reactor with an inner tube diameter of 20 mm and a total volume of approximately 380 ml. Five ports along the primary axis were used to take samples throughout the cultivation. Oxygen saturation in the PFR was monitored close to ports P1, P2 and P5 and additional aeration of 0.15 standard litres per minute was provided next to port P1 to ensure levels above 30% saturation to ambient air conditions throughout the entire PFR passage. Temperature in the PFR was maintained at 36–37°C by water heating and isolation material. A diaphragm metering pump (Sigma/1, ProMinent, Heidelberg, Germany) was used to transfer biosuspension from the stirred tank reactor to the plug flow reactor after connection of the two reactors.

Preculture, batch cultivation and continuous cultivation

A small amount of glycerol stock seed culture was spread onto 2xYT agar plates and incubated at 37°C for 24 h. A single colony was picked to inoculate 500 ml baffled shaking flasks with 50 ml of preculture minimal media. Flasks were then incubated at 37°C on an orbital shaker set to 150 rpm for 16 h. In the next morning 500 µl of biosuspension were transferred to 1000 ml baffled shaking flasks containing 100 ml preculture minimal media and incubated at 37°C on an orbital shaker set to 150 rpm for 8 h. 50 ml of this culture were used to inoculate the bioreactor. Total volume in the bioreactor was 1.6 l after inoculation. Batch fermentation in the bioreactor ensued at 28-30°C overnight. In the next morning

feed and harvest trains were connected and a constant feed/harvest rate at 5.33 ml min^{-1} corresponding to a dilution rate of 0.2 h^{-1} established. After 25 h (five volumetric residence times) of STR cultivation a reference sample was taken. The plug-flow reactor was then connected to the primary reactor via a diaphragm metering pump effectively circulating about one-quarter of the total fermentation broth from the STR through the PFR and back into the STR. In the following 28 h samples were taken at predefined time points from the STR and the five PFR ports. After 28 h of STR-PFR cultivation the fermentation was aborted, and the final broth volume measured. This value was used for all volumetric calculations during data analysis.

Determination of optical density and biomass

In preliminary experiments with identical setup correlation factors of optical density and biomass as cell dry weight (CDW) were determined for *E. coli* MG1655 and *E. coli* SR (Supporting information, Table S1). The resulting correlation factors for converting OD_{600nm} values to g l^{-1} cell dry weight were 0.324 for *E. coli* MG1655 and 0.321 for *E. coli* SR. In the main cultivations optical density was measured from appropriately diluted broth on a spectrophotometer at 600 nm and converted into biomass concentration.

Determination of acetic acid, ammonium and glucose concentrations

Five millilitres of biosuspension was directly sampled into a syringe connected to a single-use 0.45 μm sterile filter and immediately sterile filtered. The clear supernatant was flash frozen in liquid nitrogen and stored at −70°C until analysis. Glucose concentration was determined by D-Glucose UV-Test Kit (R-Biopharm, Darmstadt, Germany) and acetic acid concentration by Acetic acid UV-Test Kit (R-Biopharm, Darmstadt, Germany). Ammonium concentration was determined by Ammonium cuvette test LCK 304 (Hach Lange, Düsseldorf, Germany). At the end of the cultivation feed samples were taken and processed identically.

Analysis of total carbon, inorganic carbon and biomass composition

For total carbon and inorganic carbon analysis 0.5 ml biosuspension sample were mixed with 50 μl of 5 M KOH to prevent loss of dissolved carbonate. The suspension was then diluted 1:20 with demineralized water and stored at 4°C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) to yield the total concentration of carbon and inorganic carbon in the fermenter effluent stream. At the end of the cultivation feed samples were taken and processed identically.

To determine biomass composition 1.0 ml of biosuspension was centrifuged at 4°C and 14 000 rpm (20817 g) for 3 min. The supernatant was discarded, the pellet resuspended in 1.0 ml of freshly prepared 0.9% NaCl solution and centrifuged again. The pellet was resuspended in 5 ml 0.9% NaCl, flash frozen in liquid nitrogen and stored at −70°C until analysis. Analysis was performed with a multi N/C 2100 S composition analyzer (Analytik Jena, Jena, Germany) and the carbon content of the biomass calculated from these values.

Measurement of ppGpp

Two millilitres of of biosuspension was sampled directly into 0.5 ml of precooled (< −20°C) quenching solution and incubated at 6°C on a shaker for 15 min. Quenching solution consisted of 80 μM EDTA dissolved in 35% (V/V) perchloric acid. 500 μl 1M K_2HPO4 was added and the sample briefly vortexed. 550 μl 5 M KOH was added and the sample vortexed again. To remove precipitating potassium perchlorate samples were then centrifuged at 4°C and 7830 rpm (7197 g) for 5 min. 1.5 ml of supernatant was carefully transferred to new tubes, flash frozen in liquid nitrogen and stored at −70°C. Prior to analysis samples were thawed and their pH adjusted to 6.95 – 7.05 with 5 M KOH or 35% (V/V) perchloric acid. Samples were centrifuged again to remove all potassium perchlorate precipitate. HPLC analysis was carried out as described previously (Löffler *et al.*, 2016). If necessary, quantification was conducted by ppGpp standard addition (TriLink, San Diego, CA, USA). Samples from one time-point were analysed directly in sequence and the data normalized to the sample drawn from the STR to eliminate differences caused by column aging.

Transcriptome analysis

0.5 ml broth was sampled from the bioreactor and directly flash-frozen in liquid nitrogen. Frozen broth was then stored at −70°C until the day of RNA isolation. Total RNA was isolated using RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Isolated RNA was DNAse treated and shipped to commercial sequencing partner GENEWIZ® on dry ice. Samples were treated for rRNA depletion, sequencing libraries prepared and Illumina HiSeq 2x150 bp sequencing performed. Raw FASTQ files were obtained for bioinformatic analysis. Trimmomatic v. 0.32 (Bolger *et al.*, 2014) was used to remove adapters and low-quality reads (<Q20) checked by fastqc reports. Genes were aligned to the NCBI *E. coli* K-12 MG1655

reference genome (RefSeq: NC_000913.3) using the RNA-sequencing aligner Bowtie2 v. 2.3.2.2 (Langmead and Salzberg, 2012). On average the mapping of the reads covered 96.2%. Aligned reads were counted for each gene based on the corresponding annotation available from the NCBI database for the chosen reference sequence applying HTseq-count v. 0.6.1 in the union mode (Anders *et al.*, 2015). On average 86.4 % of the sequenced reads could be assigned uniquely to annotated features. Sequencing depth was around 27 million reads per sample on average with a mean quality phred score of 37.63.

Differential gene expression analysis was performed with the R-package DeSeq2 v. 1.26.0 (Love *et al.*, 2014) available from Bioconductor (Gentleman *et al.*, 2004). Prior to statistical analysis, all residual non-protein encoding RNA molecules (tRNA, rRNA and sRNA) were removed from the HTseq-derived raw count data and a non-specific filter was applied to remove low coverage genes with fewer than two counts per million (54 reads on average). All filtering steps caused deviations from the raw data of less than 6 %. Samples were grouped by replicates and an experimental design was chosen that used sample time and location (STR or PFR port 5) as a combined environmental factor. To normalize read counts for the comparison of sequencing depth and RNA composition, DESeq2 uses the median of ratios method to derive a scaling factor. Dividing the original read counts by the scaling factor generated normalized count values. No outliers were observed in the two biological replicates using Pearson correlation. Resulting p-values were adjusted for multiple testing according to control the false discovery rate (FDR) (Benjamini and Hochberg, 1995). Genes were identified as significantly differentially expressed by applying FDR adjusted *P*-values < 0.01 and a $\log_2$ fold change $\geq |1|$.

A principal component analysis was used to display the sample to sample distances calculated within the DESeq2 package (negative binomial distribution model). Principal component analysis was performed using plotPCA.san available on Github (https://gist.github.com/sansense/3399064897f1252d31b23ea5178c033c).

Gene set enrichment and overrepresentation analysis of up- and downregulated genes were performed using the Bioconductors's R-package GAGE v. 2.36.0 (Luo *et al.*, 2009). GAGE tests whether the mean fold-change of a gene subset is significantly different from the background using a two-tailed t-test. Genes were selected as significantly different with an FDR adjusted *P*-value < 0.01 (Benjamini and Hochberg, 1995). Functional annotation were derived from the Cluster of Orthologous Groups (COG) database (Tatusov *et al.*, 2003), the experimental sigma factor-gene interaction dataset from RegulonDB v. 10.6.3 (Santos-Zavaleta *et al.*, 2019) and the Gene Ontology (GO) Groups database with the function go.gsets from GAGE (Luo *et al.*, 2009). Furthermore, Venn diagrams were used to identify significant genes shared by both strains and differences in gene expression regulation (Chen and Boutros, 2011).

The RNA sequencing data derived from periodic ammonia starvation experiments have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO series accession number GSE158198 (Edgar *et al.*, 2002). Raw counts and processed data can be found in the Supporting information. Data analysis was performed using the free statistical computing environment R v. 3.6.2.

Acknowledgements

The authors would like to thank the group of Computational Biology at the Institute of Biochemical Engineering for the use of the Galaxy-Server.

Conflict of interest

The authors declare that they have no conflicts of interest.

Author contributions

Prof. Dr.-Ing. Ralf Takors advised the study during the entire investigation. Martin Ziegler performed the experiments, and Julia Zieringer conducted the transcriptomic analysis. Evaluation and writing of the manuscript were equally accomplished by Martin Ziegler and Julia Zieringer.

References

Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996) An Escherichia coli chromosomal "addiction module" regulated by guanosine corrected 3',5'-bispyrophosphate: a model for programmed bacterial cell death. *Proc Natl Acad Sci USA* **93:** 6059–6063. https://doi.org/10.1073/pnas.93.12.6059

Alibhai, M., and Villafranca, J.J. (1994) Kinetic and mutagenic studies of the role of the active site residues Asp-50 and Glu-327 of Escherichia coli glutamine synthetase. *Biochemistry* **33:** 682–686. https://doi.org/10.1021/bi00169a008

Allison, K.R., Brynildsen, M.P., and Collins, J.J. (2011) Heterogeneous bacterial persisters and engineering approaches to eliminate them. *Curr Opin Microbiol* **14:** 593–598. https://doi.org/10.1016/j.mib.2011.09.002.

Anders, S., Pyl, P.T., and Huber, W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)* **31:** 166–169. https://doi.org/10.1093/bioinformatics/btu638

Andrews, S.C., and Guest, J.R. (1988) Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12. *Biochem J* **255:** 35–43. https://doi.org/10.1042/bj2550035.

Ankenbauer, A., Schäfer, R.A., Viegas, S.C., Pobre, V., Voß, B., Arraiano, C.M., and Takors, R. (2020) *Pseudomonas putida* KT2440 is naturally endowed to withstand industrial-scale stress conditions. *Microb Biotechnol* **13:** 1145–1161. https://doi.org/10.1111/1751-7915.13571

Atherly, A.G. (1979) *Escherichia coli* mutant containing a large deletion from relA to argA. *J. Bacteriol.* **138:** 530–534.

Balaban, N.Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004) Bacterial persistence as a phenotypic switch. *Science* **305:** 1622–1625. https://doi.org/10.1126/science.1099390

Barker, M.M., Gaal, T., Josaitis, C.A., and Gourse, R.L. (2001) Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J Mol Biol* **305:** 673–688. https://doi.org/10.1006/jmbi.2000.4327

Barker, C.S., Prüss, B.M., and Matsumura, P. (2004) Increased motility of *Escherichia coli* by insertion sequence element integration into the regulatory region of the flhD operon. *J. Bacteriol.* **186:** 7529–7537. https://doi.org/10.1128/JB.186.22.7529-7537.2004

Bekker, M., de Vries, S., Ter Beek, A., Hellingwerf, K.J., and de Mattos, M.J.T. (2009) Respiration of *Escherichia coli* can be fully uncoupled via the nonelectrogenic terminal cytochrome bd-II oxidase. *J Bacteriol* **191:** 5510–5517. https://doi.org/10.1128/JB.00562-09

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc: Ser B (Methodol)* **57:** 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bi, E., and Lutkenhaus, J. (1993) Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring. *J Bacteriol* **175:** 1118–1125. https://doi.org/10.1128/jb.175.4.1118-1125.1993

Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* **30:** 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Brown, D.R., Barton, G., Pan, Z., Buck, M., and Wigneshweraraj, S. (2014) Nitrogen stress response and stringent response are coupled in *Escherichia coli*. *Nat Commun* **5:** 4115. https://doi.org/10.1038/ncomms5115

Butala, M., Klose, D., Hodnik, V., Rems, A., Podlesek, Z., Klare, J.P., *et al.* (2011) Interconversion between bound and free conformations of LexA orchestrates the bacterial SOS response. *Nucleic Acids Res* **39:** 6546–6557. https://doi.org/10.1093/nar/gkr265

Bylund, F., Castan, A., Mikkola, R., Veide, A., and Larsson, G. (2000) Influence of scale-up on the quality of recombinant human growth hormone. *Biotechnol Bioeng* **69:** 119–128. https://doi.org/10.1002/(SICI)1097-0290(20000720)69:2<119:AID-BIT1>3.0.CO;2-9

Bylund, F., Collet, E., Enfors, S.-O., and Larsson, G. (1998) Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Eng* **18:** 171. https://doi.org/10.1007/s004490050427

Bylund, F., Guillard, F., Enfors, S.-O., Trägårdh, C., and Larsson, G. (1999) Scale down of recombinant protein production: a comparative study of scaling performance. *Bioprocess Eng* **20:** 377. https://doi.org/10.1007/s004490050606

Chang, D.-E., Smalley, D.J., and Conway, T. (2002) Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol Microbiol* **45:** 289–306. https://doi.org/10.1046/j.1365-2958.2002.03001.x

Chen, H., and Boutros, P.C. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12:** 35. https://doi.org/10.1186/1471-2105-12-35

Chowdhury, N., Kwan, B.W., and Wood, T.K. (2016) Persistence increases in the absence of the alarmone guanosine tetraphosphate by reducing cell growth. *Sci Rep* **6:** 20519. https://doi.org/10.1038/srep20519

Chubukov, V., Desmarais, J.J., Wang, G., Chan, L.J.G., Baidoo, E.E., Petzold, C.J., *et al.* (2017) Engineering glucose metabolism of *Escherichia coli* under nitrogen starvation. *NPJ Syst Biol App* **3:** 16035. https://doi.org/10.1038/npjsba.2016.35

Dalebroux, Z.D., and Swanson, M.S. (2012) ppGpp: magic beyond RNA polymerase. *Nat Rev Microbiol* **10:** 203–212. https://doi.org/10.1038/nrmicro2720

Delvigne, F., Destain, J., and Thonart, P. (2006) A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models. *Biochem Eng J* **28:** 256–268. https://doi.org/10.1016/j.bej.2005.11.009

Delvigne, F., Takors, R., Mudde, R., van Gulik, W., and Noorman, H. (2017) Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. *Microb Biotechnol* **10:** 1267–1274. https://doi.org/10.1111/1751-7915.12803

Doucette, C.D., Schwab, D.J., Wingreen, N.S., and Rabinowitz, J.D. (2011) α-Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition. *Nat Chem Biol* **7:** 894–901. https://doi.org/10.1038/nchembio.685

Durfee, T., Hansen, A.-M., Zhi, H., Blattner, F.R., and Jin, D.J. (2008) Transcription profiling of the stringent response in *Escherichia coli*. *J Bacteriol* **190:** 1084–1096. https://doi.org/10.1128/JB.01092-07

Edgar, R., Domrachev, M., and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210. https://doi.org/10.1093/nar/30.1.207

Enfors, S.-O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., *et al.* (2001) Physiological responses to mixing in large scale bioreactors. *J Biotechnol* **85:** 175–185. https://doi.org/10.1016/S0168-1656(00)00365-5

Figueira, R., Brown, D.R., Ferreira, D., Eldridge, M.J.G., Burchell, L., Pan, Z., *et al.* (2015) Adaptation to sustained nitrogen starvation by *Escherichia coli* requires the eukaryote-like serine/threonine kinase YeaG. *Sci Rep* **5:** 17524. https://doi.org/10.1038/srep17524

Fonville, N.C., Bates, D., Hastings, P.J., Hanawalt, P.C., and Rosenberg, S.M. (2010) Role of RecA and the SOS response in thymineless death in *Escherichia coli*. *PLoS*

Genet **6:** e1000865. https://doi.org/10.1371/journal.pgen.1000865

Gaca, A.O., Colomer-Winter, C., and Lemos, J.A. (2015) Many means to a common end: the intricacies of (p)ppGpp metabolism and its control of bacterial homeostasis. *J Bacteriol* **197:** 1146–1156. https://doi.org/10.1128/JB.02577-14

Gallant, J., Erlich, H., Hall, B., and Laffler, T. (1970) Analysis of the RC function. *Cold Spring Harb Symp Quant Biol* **35:** 397–405. https://doi.org/10.1101/SQB.1970.035.01.051

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5:** R80. https://doi.org/10.1186/gb-2004-5-10-r80

George, S., Larsson, G., and Enfors, S.-O. (1993) A scale-down two-compartment reactor with controlled substrate oscillations: metabolic response of *Saccharomyces cerevisiae*. *Bioprocess Eng* **9:** 249–257. https://doi.org/10.1007/BF01061530

Gerdes, K., and Maisonneuve, E. (2012) Bacterial persistence and toxin-antitoxin loci. *Annu Rev Microbiol* **66:** 103–123. https://doi.org/10.1146/annurev-micro-092611-150159

Gerding, M.A., Ogata, Y., Pecora, N.D., Niki, H., and de Boer, P.A.J. (2007) The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. *Mol Microbiol* **63:** 1008–1025. https://doi.org/10.1111/j.1365-2958.2006.05571.x

González Barrios, A.F., Zuo, R., Hashimoto, Y., Yang, L., Bentley, W.E., and Wood, T.K. (2006) Autoinducer 2 controls biofilm formation in *Escherichia coli* through a novel motility quorum-sensing regulator (MqsR, B3022). *J Bacteriol* **188:** 305–316. https://doi.org/10.1128/JB.188.1.305-316.2006

Hardiman, T., Lemuth, K., Keller, M.A., Reuss, M., and Siemann-Herzberg, M. (2007) Topology of the global regulatory network of carbon limitation in *Escherichia coli*. *J Biotechnol* **132:** 359–374. https://doi.org/10.1016/j.jbiotec.2007.08.029

Haringa, C., Deshmukh, A.T., Mudde, R.F., and Noorman, H.J. (2017) Euler-Lagrange analysis towards representative down-scaling of a 22 m 3 aerobic *S. cerevisiae* fermentation. *Chem Eng Sci* **170:** 653–669. https://doi.org/10.1016/j.ces.2017.01.014

Hauryliuk, V., Atkinson, G.C., Murakami, K.S., Tenson, T., and Gerdes, K. (2015) Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* **13:** 298–309. https://doi.org/10.1038/nrmicro3448

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., *et al.* (2006) Highly accurate genome sequences of *Escherichia coli* K-12 strains MG1655 and W3110. *Mol Syst Biol* **2,** 2006.007. https://doi.org/10.1038/msb4100049

Hopkins, D.J., Betenbaugh, M.J., and Dhurjati, P. (1987) Effects of dissolved oxygen shock on the stability of recombinant *Escherichia coli* containing plasmid pKN401. *Biotechnol Bioeng* **29:** 85–91. https://doi.org/10.1002/bit.260290113

Hua, Q., Yang, C., Oshima, T., Mori, H., and Shimizu, K. (2004) Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. *Appl Environ Microbiol* **70:** 2354–2366. https://doi.org/10.1128/AEM.70.4.2354-2366.2004

Huisman, O., D'Ari, R., and Gottesman, S. (1984) Cell-division control in *Escherichia coli*: specific induction of the SOS function SfiA protein is sufficient to block septation. *Proc Natl Acad Sci USA* **81:** 4490–4494. https://doi.org/10.1073/pnas.81.14.4490

Jarmander, J., Belotserkovsky, J., Sjöberg, G., Guevara-Martínez, M., Pérez-Zabaleta, M., Quillaguamán, J., and Larsson, G. (2015) Cultivation strategies for production of (R)-3-hydroxybutyric acid from simultaneous consumption of glucose, xylose and arabinose by *Escherichia coli*. *Microb Cell Fact* **14:** 51. https://doi.org/10.1186/s12934-015-0236-2.

Jishage, M., Iwata, A., Ueda, S., and Ishihama, A. (1996) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *J Bacteriol* **178:** 5447–5451. https://doi.org/10.1128/jb.178.18.5447-5451.1996

Johansson, J., Balsalobre, C., Wang, S.-Y., Urbonaviciene, J., Jin, D.J., Sondén, B., and Uhlin, B.E. (2000) Nucleoid proteins stimulate stringently controlled bacterial promoters. *Cell* **102:** 475–485. https://doi.org/10.1016/S0092-8674(00)00052-0

de Jonge, L.P., Buijs, N.A.A., ten Pierick, A., Deshmukh, A., Zhao, Z., Kiel, J.A.K.W., *et al.* (2011) Scale-down of penicillin production in *Penicillium chrysogenum*. *Biotechnol J* **6:** 944–958. https://doi.org/10.1002/biot.201000409

Joseleau-Petit, D., Vinella, D., and D'Ari, R. (1999) Metabolic alarms and cell division in *Escherichia coli*. *J Bacteriol* **181:** 9–14. https://doi.org/10.1128/JB.181.1.9-14.1999

Junker, B.H. (2004) Scale-up methodologies for *Escherichia coli* and yeast fermentation processes. *J Biosci Bioeng* **97:** 347–364. https://doi.org/10.1016/S1389-1723(04)70218-2

Junne, S., Klingner, A., Kabisch, J., Schweder, T., and Neubauer, P. (2011) A two-compartment bioreactor system made of commercial parts for bioprocess scale-down studies: impact of oscillations on *Bacillus subtilis* fed-batch cultivations. *Biotechnol J* **6:** 1009–1017. https://doi.org/10.1002/biot.201100293

Kanjee, U., Gutsche, I., Alexopoulos, E., Zhao, B., El Bakkouri, M., Thibault, G., *et al.* (2011) Linkage between the bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase. *EMBO J* **30** 931–944. https://doi.org/10.1038/emboj.2011.5

Kanjee, U., Ogata, K., and Houry, W.A. (2012) Direct binding targets of the stringent response alarmone (p)ppGpp. *Mol Microbiol* **85:** 1029–1043. https://doi.org/10.1111/j.1365-2958.2012.08177.x

Kelly, W.J. (2008) Using computational fluid dynamics to characterize and improve bioreactor performance. *Biotechnol Appl Biochem* **49(Pt** 4): 225–238. https://doi.org/10.1042/BA20070177

Korch, S.B., Malhotra, V., Contreras, H., and Clark-Curtiss, J.E. (2015) The *Mycobacterium tuberculosis* relBE toxin: antitoxin genes are stress-responsive modules that

regulate growth through translation inhibition. *J Microbiol (Seoul, Korea)* **53(11):** 783–795. https://doi.org/10.1007/s12275-015-5333-8

Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. https://doi.org/10.1038/nmeth.1923

Lapin, A., Schmid, J., and Reuss, M. (2006) Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor—A structured—segregated approach. *Chem Eng Sci* **61:4783–4797.** https://doi.org/10.1016/j.ces.2006.03.003

Lara, A.R., Galindo, E., Ramírez, O.T., and Palomares, L.A. (2006) Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. *Mol Biotechnol* **34:** 355–382. https://doi.org/10.1385/MB:34:3:355

Lara, A.R., Taymaz-Nikerel, H., Mashego, M.R., van Gulik, W.M., Heijnen, J.J., Ramírez, O.T., and van Winden, W.A. (2009) Fast dynamic response of the fermentative metabolism of *Escherichia coli* to aerobic and anaerobic glucose pulses. *Biotechnol Bioeng* **104:** 1153–1161. https://doi.org/10.1002/bit.22503

Larsson, G., and Enfors, S.-O. (1988) Studies of insufficient mixing in bioreactors: effects of limiting oxygen concentrations and short term oxygen starvation on *Penicillium chrysogenum*. *Bioprocess Eng* **3:** 123–127. https://doi.org/10.1007/BF00373475

Larsson, G., Trnkvist, M., Wernersson, E.S., Trgrdh, C., Noorman, H., and Enfors, S.-O. (1996) Substrate gradients in bioreactors: origin and consequences. *Bioprocess Eng* **14:** 281–289. https://doi.org/10.1007/BF00369471

Lee, S., Kim, M.H., Kang, B.S., Kim, J.-S., Kim, G.-H., Kim, Y.-G., and Kim, K.J. (2008) Crystal structure of *Escherichia coli* MazG, the regulator of nutritional stress response. *J Biol Chem* **283:** 15232–15240. https://doi.org/10.1074/jbc.M800479200

Löffler, M., Simen, J.D., Jager, G., Schaferhoff, K., Freund, A., and Takors, R. (2016) Engineering *E. coli* for large-scale production - Strategies considering ATP expenses and transcriptional responses. *Metab Eng* **38:** 73–85. https://doi.org/10.1016/j.ymben.2016.06.008

Love, M.I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15:** 550. https://doi.org/10.1186/s13059-014-0550-8

Luo, W., Friedman, M.S., Shedden, K., Hankenson, K.D., and Woolf, P.J. (2009) GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinform* **10:** 161. https://doi.org/10.1186/1471-2105-10-161

Magnusson, L.U., Farewell, A., and Nyström, T. (2005) ppGpp: a global regulator in *Escherichia coli*. *Trends Microbiol* **13:** 236–242. https://doi.org/10.1016/j.tim.2005.03.008

Michalowski, A., Siemann-Herzberg, M., and Takors, R. (2017) Escherichia coli HGT: engineered for high glucose throughput even under slowly growing or resting conditions. *Metab Eng* **40:** 93–103. https://doi.org/10.1016/j.ymben.2017.01.005

Mueller, E.J., Oh, S., Kavalerchik, E., Kappock, T.J., Meyer, E., Li, C., *et al.* (1999) Investigation of the ATP binding site of *Escherichia coli* aminoimidazole ribonucleotide synthetase using affinity labeling and site-directed mutagenesis. *Biochemistry* **38:** 9831–9839. https://doi.org/10.1021/bi990638r

Murray, K.D., and Bremer, H. (1996) Control of spoT-dependent ppGpp synthesis and degradation in *Escherichia coli*. *J Mol Biol* **259:** 41–57.

Neubauer, P., Åhman, M., Törnkvist, M., Larsson, G., and Enfors, S.-O. (1995a) Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of *Escherichia coli*. *J Biotechnol* **43:** 195–204. https://doi.org/10.1016/0168-1656(95)00130-1

Neubauer, P., Häggström, L., and Enfors, S.O. (1995b) Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations. *Biotechnol Bioeng* **47:** 139–146. https://doi.org/10.1002/bit.260470204

Neubauer, P., and Junne, S. (2010) Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Curr Opin Biotechnol* **21:** 114–121. https://doi.org/10.1016/j.copbio.2010.02.001

Noguchi, Y., Nakai, Y., Shimba, N., Toyosaki, H., Kawahara, Y., Sugimoto, S., and Suzuki, E.-I. (2004) The energetic conversion competence of *Escherichia coli* during aerobic respiration studied by 31P NMR using a circulating fermentation system. *J Biochem* **136:** 509–515. https://doi.org/10.1093/jb/mvh147

Noorman, H. (2011) An industrial perspective on bioreactor scale-down: what we can learn from combined large-scale bioprocess and model fluid studies. *Biotechnol J* **6:** 934–943. https://doi.org/10.1002/biot.201000406

Oliveira-Filho, E.R., Silva, J.G.P., de Macedo, M.A., Taciro, M.K., Gomez, J.G.C., and Silva, L.F. (2019) Investigating nutrient limitation role on improvement of growth and Poly (3-Hydroxybutyrate) accumulation by *Burkholderia sacchari* LMG 19450 from xylose as the sole carbon source. *Front Bioeng Biotechnol* **7:** 416. https://doi.org/10.3389/fbioe.2019.00416

Perez-Zabaleta, M., Sjöberg, G., Guevara-Martínez, M., Jarmander, J., Gustavsson, M., Quillaguamán, J., and Larsson, G. (2016) Increasing the production of (R)-3-hydroxybutyrate in recombinant *Escherichia coli* by improved cofactor supply. *Microb Cell Fact* **15:** 91. https://doi.org/10.1186/s12934-016-0490-y

Reitzer, L. (2003) Nitrogen assimilation and global regulation in escherichia coli. *Annu Rev Microbiol* **57:** 155–176. https://doi.org/10.1146/annurev.micro.57.030502.090820

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeida, D., *et al.* (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res* **47:** D212–D220. https://doi.org/10.1093/nar/gky1077

Simen, J.D., Löffler, M., Jäger, G., Schäferhoff, K., Freund, A., Matthes, J., *et al.* (2017) Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. *Microb Biotechnol* **10:** 858–872. https://doi.org/10.1111/1751-7915.12713

Sun, C., Guo, Y., Tang, K., Wen, Z., Li, B., Zeng, Z., and Wang, X. (2017) MqsR/MqsA toxin/antitoxin system regulates persistence and biofilm formation in *Pseudomonas*

putida KT2440. *Front Microbiol* **8:** 840. https://doi.org/10.3389/fmicb.2017.00840

Szenk, M., Dill, K.A., and de Graff, A.M.R. (2017) Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis. *Cell systems* **5:** 95–104. https://doi.org/10.1016/j.cels.2017.06.005

Takors, R. (2012) Scale-up of microbial processes: impacts, tools and open questions. *J Biotechnol* **160:** 3–9. https://doi.org/10.1016/j.jbiotec.2011.12.010

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinform* **4:** 41. https://doi.org/10.1186/1471-2105-4-41

Traxler, M.F., Summers, S.M., Nguyen, H.-T., Zacharia, V.M., Hightower, G.A., Smith, J.T., and Conway, T. (2008) The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol Microbiol* **68:** 1128–1148. https://doi.org/10.1111/j.1365-2958.2008.06229.x

Traxler, M.F., Zacharia, V.M., Marquardt, S., Summers, S.M., Nguyen, H.-T., Stark, S.E., and Conway, T. (2011) Discretely calibrated regulatory loops controlled by ppGpp partition gene induction across the 'feast to famine' gradient in *Escherichia coli*. *Mol Microbiol* **79:** 830–845. https://doi.org/10.1111/j.1365-2958.2010.07498.x

Vasilakou, E., van Loosdrecht, M.C.M., and Wahl, S.A. (2020) *Escherichia coli* metabolism under short-term repetitive substrate dynamics: adaptation and trade-offs. *Microb Cell Fact* **19:** 116. https://doi.org/10.1186/s12934-020-01379-0

Vrábel, P., van der Lans, R.G., Luyben, K.C., Boon, L., and Nienow, A.W. (2000) Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: modelling and measurements. *Chem Eng Sci* **55:** 5881–5896. https://doi.org/10.1016/S0009-2509(00)00175-5

Wang, J.D., and Levin, P.A. (2009) Metabolism, cell growth and the bacterial cell cycle. *Nat Rev Microbiol* **7:** 822–827. https://doi.org/10.1038/nrmicro2202

Wang, Q., Yu, H., Xia, Y., Kang, Z., and Qi, Q. (2009) Complete PHB mobilization in *Escherichia coli* enhances the stress tolerance: a potential biotechnological application. *Microb Cell Fact* **8:** 47. https://doi.org/10.1186/1475-2859-8-47

Wen, Q., Chen, Z., Tian, T., and Chen, W. (2010) Effects of phosphorus and nitrogen limitation on PHA production in activated sludge. *J Environ Sci* **22:** 1602–1607. https://doi.org/10.1016/S1001-0742(09)60295-3

Williamson, G., Tamburrino, G., Bizior, A., Boeckstaens, M., Dias Mirandela, G., Bage, M., et al. (2020) A two-lane mechanism for selective biological ammonium transport. *eLife* **9:** https://doi.org/10.7554/eLife.57183

Xi, H., Schneider, B.L., and Reitzer, L. (2000) Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J Bacteriol* **182:** 5332–5341. https://doi.org/10.1128/JB.182.19.5332-5341.2000

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Long term changes STR 5min vs STR 0h.
Table S2. Long term changes STR 28h vs STR 0h.
Table S3. Short term changes PFR28h vs STR28h.
Table S4. Short term changes PFR5min vs STR5min.
Appendix S1. Supplementary Information.

Received: 17 July 2020 | Revised: 11 September 2020 | Accepted: 11 September 2020

DOI: 10.1002/bit.27568

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Data-driven in silico prediction of regulation heterogeneity and ATP demands of *Escherichia coli* in large-scale bioreactors

Julia Zieringer | Moritz Wild | Ralf Takors

Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, Germany

Correspondence
Ralf Takors, Institute of Biochemical Engineering, University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany.
Email: takors@ibvt.uni-stuttgart.de

Abstract

Escherichia coli exposed to industrial-scale heterogeneous mixing conditions respond to external stress by initiating short-term metabolic and long-term strategic transcriptional programs. In native habitats, long-term strategies allow survival in severe stress but are of limited use in large bioreactors, where microenvironmental conditions may change right after said programs are started. Related on/off switching of genes causes additional ATP burden that may reduce the cellular capacity for producing the desired product. Here, we present an agent-based data-driven model linked to computational fluid dynamics, finally allowing to predict additional ATP needs of *Escherichia coli* K12 W3110 exposed to realistic large-scale bioreactor conditions. The complex model describes transcriptional up- and downregulation dynamics of about 600 genes starting from subminute range covering 28 h. The data-based approach was extracted from comprehensive scale-down experiments. Simulating mixing and mass transfer conditions in a 54 m^3 stirred bioreactor, 120,000 *E. coli* cells were tracked while fluctuating between different zones of glucose availability. It was found that cellular ATP demands rise between 30% and 45% of growth decoupled maintenance needs, which may limit the production of ATP-intensive product formation accordingly. Furthermore, spatial analysis of individual cell transcriptional patterns reveal very heterogeneous gene amplifications with hot spots of 50%–80% messenger RNA upregulation in the upper region of the bioreactor. The phenomenon reflects the time-delayed regulatory response of the

NOMENCLATURE: *a*, *b*, *c*, *d*, regression paramete; *C*, cluster; c_S, glucose concentration (mol L^{-1}); *D*, dilution rate (h^{-1}); *Deg*, degradation; *f*, fading genes; *gene*, gene; *i*, iteration index; k_{deg}, protein degradation rate (h^{-1}); K_S, substrate-specific uptake constant (g L^{-1}); $L_{mRNA,i}$, length of mRNA strand per gene *i* (nt); max_{TL}, maximum number of translations per mRNA; *Min size*, minimum number of observations per clusters; *MW*, molecular weight (g mol^{-1}); *Ntry*, number of trials per iteration; $n_{nucleotide}$, number of nucleotide monophosphates; *p*, persisting genes; *protein*, bulk protein; $q_{s,\ max}$, maximum substrate uptake rate (g(gh)$^{-1}$); *rmax*, maximum radius as a proxy for correlation; *S*, number of starvation passages; *Sc*, Schmidt number; S_{crit}, critical number of starvation passages; S_{min}, minimum number of starvation passages; *t*, time (general) (s); $t_{deg,\ med}$, median lifetime of mRNA (s); $TL_{i,\ j}$, initiation trigger for a translation j on mRNA *i*; v_{RIB}, velocity RIB (nts^{-1}); v_{RNAP}, velocity RNAP (nts^{-1}); x_{RIB}, location of RIB on mRNA (nt); x_{RNAP}, location of RNAP on DNA (nt); Δx_{RIB}, distance between ribosomes (nt); Δx_{RNAP}, distance between RNAPs (nt); Y_{XS}, substrate-biomass Yield ($g_s g^{-1}_{CDW}$); $\sigma_{RNAP}(t)$, number of active RNAPs; τ, residence time (s); μ, growth rate (h^{-1}).

cells that propagate through the stirred tank. After 4.2 h, cells adapt to environmental changes but still have to bear an additional 6% ATP demand.

KEYWORDS
agent-based modeling, ATP maintenance, computational fluid dynamics (CFD), *Escherichia coli*, glucose gradient, heterogeneities

1 | INTRODUCTION

To reduce the human CO_2 footprint in the atmosphere, sustainable bioprocesses replacing fossil resources by sugar may play a crucial role. Microbial production offers the potential to provide products for agricultural, biopharmaceutical, and chemical markets (Delvigne et al., 2017). As a prerequisite, such approaches need to be transferred successfully from laboratory to large-scale without loss of economic attraction, that is, without reduction of the sensitive TRY values (titer, rates, and yields) that served as constraints for economic evaluation. However, performance losses may occur, comprising increased by-product formation, reduced substrate-to-product conversion, reduced productivities, and so forth (Lara et al., 2006). They mirror cellular responses to large-scale heterogeneities that are induced by limited mass transfer and by insufficient mixing capacities (Noorman & Heijnen, 2017). Accordingly, research activities aimed to mimic large-scale conditions already in early-stage lab tests. One of the first examples is given by Oosterhuis and Kossen (1983), who repeatedly exposed cells to oxygen saturated and limiting conditions in a setting of two linked, stirred bioreactors. Multiple tests with alternate experimental scale-up simulators followed (overviews provided in Garcia-Ochoa & Gomez, 2009; Neubauer & Junne, 2010, 2016; Noorman, 2011; Takors, 2012; Zieringer & Takors, 2018) mimicking not only fluctuations of dissolved oxygen (DO) levels, but also nutrient availability and pH variations. Today, such approaches received key consideration to design robust microbial processes (Noorman & Heijnen, 2017). Still, the valid a priori prediction of large-scale heterogeneities' impact on cellular performance is of crucial importance for developing novel bioprocesses. Even further, findings of large-scale stress exposure may guide strain engineering to create particularly robust hosts. To reach this goal, Löffler et al. (2016) applied the so-called STR/PFR setup comprising a stirred tank reactor (STR) linked with a plug-flow reactor (PFR). Steady-state nutrient-limited, continuous cultivations were performed in STR before PFR was connected, frequently exposing cells to glucose-limiting conditions. Accordingly, cells repeatedly experienced temporal feast/famine conditions that were characterized by the residence time in the PFR. Comprehensive sampling in STR and PFR created a highly valuable data set of short- and long-term metabolic and transcriptional responses on repeated starvation stimuli (Löffler et al., 2016). The data set revealed that *Escherichia coli* not only react on extracellular stress by instantaneous metabolic shifts. Observations also revealed massive transcription of genes organized in operons (Nieß et al., 2017) and in fundamental regulons of strategic importance. For instance, the stringent response was repeatedly initiated by fast-rising intracellular (p)ppGpp levels in PFR, which were downregulated in STR. Löffler et al. (2016) reported an additional rise of growth-decoupled maintenance of up to 50%. So far, these findings were not yet used to predict the response of *E. coli* exposed to large-scale heterogeneities. First, a data-driven model is needed that describes the complex transcriptional response of *E. coli* to said stress conditions. Next, such a transcriptional model should be coupled with computational fluid dynamics (CFD) of a large-scale bioreactor to identify zones of different nutrient availability and to predict the cellular response of cells passing through those zones (Zieringer & Takors, 2018). Our study exactly tackles this two-step problem: Mixing heterogeneities and zones of different substrate availability of a 54 m^3 stirred bioreactor are predicted using CFD and assuming common operating conditions. The tracking of 120,000 *E. coli* cells finally yielded the prediction of additional ATP demands. Furthermore, spatially resolved transcriptional patterns of individual E. coli cells were predicted, unraveling the population heterogeneity in the industrial-scale bioreactor.

2 | MATERIALS AND METHODS

2.1 | Experimental setup

A glucose gradient was simulated in a stirred tank reactor (STR) coupled to a plug flow reactor (PFR), as depicted in Figure 1.

The experimental setup consists of an STR operated in continuous mode (dilution rate, $D = 0.2\ h^{-1}$) and connected with a plug flow reactor. Cells were grown under glucose-limited conditions in the STR (mean residence time of the cells in STR: 6.2 min) and experience starvation in the PFR (mean residence time of the cells in PFR: 125 s). The cells are circulating through the reactor system for 28 h process time, which equals, on average, around 200 passages of the starvation zone for each cell. In this way, the setup permits the analysis of transcriptional response for ongoing starvation passages through the PFR. Thereby, the tactical response is monitored via the PFR sample ports (P1–P5), while the strategic changes were tracked via the STR sample port (S). The cultivation was performed as biological triplicates under identical experimental conditions. For transcriptomic analysis, the samples were grouped by replicates, and sample time and location (STR or PFR) was chosen as a combined experimental design. Significantly expressed genes were determined using the described design and a generalized linear model within the

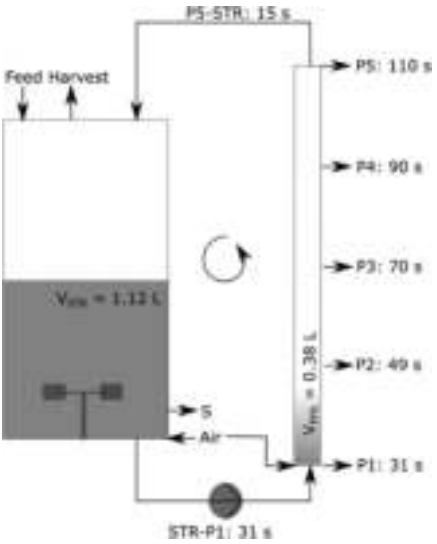


FIGURE 1 Scheme of a two-compartment system as used in the experimental setup (Löffler et al., 2016). The two-compartment device consists of a stirred tank reactor (STR) connected to a plug-flow reactor (PFR). Derived from non-ideally mixed large-scale industrial fermenters (Lapin et al., 2006), the setup mimics periodic substrate availability experienced by cells in large-scale bioreactors. The well-mixed STR is operated in glucose-limited continuous mode (dilution rate, $D = 0.2\ h^{-1}$). As soon as cells enter the PFR compartment, the residual substrate is consumed within seconds leading to starvation. The steady state before PFR onset at time zero was used as the reference state (S_0). Samples were taken at eleven distinct time points over 28 h. The system is equipped with five PFR sample ports (P1–P5) at defined residence times τ(s), as well as an STR sample port S. The total mean PFR residence time is $\tau_{PFR} = 125$ s

egdeR *R*-package (v.3.8.6) (Robinson et al., 2010). The detailed experimental implementation and RNA sequencing results used in this publication were published in the paper of Löffler et al. (2016).

2.2 | RNA-sequencing data cluster analysis

RNA-sequencing data contain time courses of messenger RNA (mRNA) abundance of 3908 genes. Thereof, three measurement sets sampled after 25 min, 2 h, and 28 h were chosen for further investigations. Significantly up- und downregulated genes of samples P1–P5 in PFR passing the threshold of *log2* fold change (log2FC) > |0.58| and false discovery rate (FDR)-corrected *p* value < .05 were identified. Cluster analysis was performed using the *R*-package *flexclust* v. 1.4-0 (Leisch, 2006) applying *RStudio* v. 1.2.1335 (RStudio, Inc.) RNA-sequencing data contain time courses of mRNA abundance of 3908 genes. Thereof, three measurement sets sampled after 25 min, 2 h, and 28 h were chosen for further investigations. Significantly up- und downregulated genes of samples P1–P5 in PFR passing the threshold of *log2* fold change (log2FC) > |0.58| and FDR-corrected *p* value < .05 were identified. Cluster analysis was performed using the *R*-package *flexclust* v. 1.4-0 (Leisch, 2006), applying *RStudio* v. 1.2.1335 to significantly reduce simulation efforts while including basic features of gene dynamics. The function *qtclust* (included in flexclust package) was used to perform stochastic quality-based clustering (SQBC) and *k*-means-clustering. Parameters of SQBC were set as follows (Table 1).

Ntry indicates the number of trials per iteration, while *rmax* is the maximum radius as a proxy for correlation. *Min size* defines the minimum number of observations per cluster. Data points not clustered by the algorithm are omitted. The setting of parameters ensured maximum comparability and five as the maximum number of clusters. Cluster properties are listed in the Supporting Information Appendix Tables D5–D8, and resulting clusters are displayed in Supporting Information C in the Appendix. The *k*-means algorithm was initialized with the centroids of the SQCB method.

2.3 | ATP calculation for single molecules

ATP requirements for the formation of amino acids and nucleotides were calculated using the results of Kaleta et al. (2013). The translational costs for protein formation and polymerization add up to 4 ATP per amino acid, including activation of the amino acid (1 ATP to 1 AMP) and peptide bond formation at the ribosome (2 GTP) (Stouthamer, 1973). Since there is a net production of 0.1 ATP per amino acid (for detailed calculation, see Löffler et al., 2016), the overall cost of amino acid synthesis and polymerization was estimated as four ATPs consumed per residue. The absolute numbers of synthesized and degraded nucleotides (*nts*) were estimated from experimental data. To recycle mono phosphorylated nucleotides (NMPs) to triphosphorylated nucleotides (NTPs), costs of 2 ATP were assumed. Assuming a P/O-ratio of 1.49 (ATP formation via NADH oxidation in respiration), the following ATP requirements were assumed for the bases (Table 2).

The growth-independent maintenance was used as 0.0027 mol $(g_{DW}h)^{-1}$, according to Taymaz-Nikerel et al. (2010).

2.3.1 | Calculation of mRNA abundance

Only additional ATP needs for transcription and translation were estimated considering the basic demands under nonperturbed conditions. Accordingly, total mRNA content was estimated following studies of Bremer and Dennis (2008) as 61.7 µg per 10^9 cells for a growth rate of $0.2\ h^{-1}$. 20% of the total dry weight was assumed to be RNA (Neidhardt et al., 1990), including 5% mRNA, a value which is

in line with conclusions from Stouthamer (1973). This results in total mRNA content of 6.17 g per 10^{16} mRNA per cell. The relative distribution of specific mRNAs is taken from the measured normalized counts as transcripts per million. The relative fraction multiplied with the total mRNA content gives the total mass of all mRNA encoded by a single gene. Dividing this number with the corresponding molecular weight yields the absolute number of molecules. Molecular weights of mRNAs were calculated with Equation 1 (Kibbe, 2007). Results are listed in Supporting Information Table A3. As the phosphate groups of two nucleotides are bound together, an OH-group is cleaved. This leads to reduced molecular weights of nucleotides in the polymer chain. Accordingly, an additional term was added to account for the 5'-triphosphate cleavage (159 g mol^{-1}).

$$MW_{mRNA} = n_{Guanine} \times 345.2 + n_{Cytosine} \times 305.2 + n_{Adenine} \times 329.2 + n_{Uracil} \times 306.2 + 159 \ (gmol^{-1}). \quad (1)$$

$n_{nucleotide}$ codes for the number of nucleotide monophosphates which is multiplied by their corresponding molecular weight.

2.4 | The biological model

The model was implemented using MATLAB v. 2019b, considering the four processes: transcription, translation, mRNA degradation, and protein degradation. The first three are implemented as agent-based approaches, while the last considers protein degradation as a decomposition in the continuum. The governing variable that controls the expression is the number of active RNA polymerases (RNAPs) per each cluster.

2.4.1 | Estimating the number of active RNAP

We assumed that gene expression levels follow sigmoidal courses. Hence, an equilibrium between synthesis and degradation may be achieved. Produced *nts* are given by

$$v_{RNAP} \cdot \int_0^t \sigma_{RNAP}(t)dt = Nt_{prod} \quad (2)$$

with $\sigma_{RNAP}(t)$ coding for the number of active RNAPs at time *t*. The shape of the sigmoidal function is defined as

$$\sigma(t) = \frac{a}{1 + e^{\{b \cdot (t+c)\}}} + d. \quad (3)$$

The parameters *a*, *b*, *c*, and *d* were fitted to the nucleotide synthesis, which was derived from experimental data by calculating the number of synthesized copies and by considering individual gene lengths. Consequently, steadily rising functions were obtained that allowed to estimate the number of active RNAPs per cluster. For the latter, a constant RNAP transcription velocity of 21 nucleotides per second was assumed (Chen et al., 2015).

2.4.2 | Transcription

After initiation, the continuous one-stranded movement of RNAP creates the mRNA transcripts measured. However, individual gene expression profiles were observed that could be grouped in clusters of similar transcription dynamics. Accordingly, only expression dynamics of representative, average genes per cluster are described in the model (Supporting Information Appendix Tables D5–D7). The minimum distance of 100 nts (Δx_{RNAP}) was considered between two subsequent RNAPs. Furthermore, all genes of one cluster were supposed to be randomly initiated with the functional

$$gene_{i,on|off}(t) = \begin{cases} 1 & \text{if free RNAP is avaliable and gene}_i \\ & \text{is randomly chosen,} \\ 0 & \text{else.} \end{cases} \quad (4)$$

Gene transcription is modeled as one-dimensional nucleotide extension with the relative movement of RNAP as

$$\frac{dx_{RNAP}}{dt} = \begin{cases} v_{RNAP} & \text{if initiated,} \\ 0 & \text{else.} \end{cases} \quad (5)$$

The constant transcriptional elongation rate, v_{RNAP}, is set to 21 nts^{-1}, which equals the average value found in *E. coli* during starvation (Chen et al., 2015). The variable x_{RNAP} indicates the relative position of RNAP on the DNA grid. The length of the resulting mRNA strand is equivalent as

$$x_{i,RNAP} = L_{i,mRNA}. \quad (6)$$

When the last nucleotide is reached, the mRNA is released. All fragments of mRNA are summed to get total mRNA amounts. While fractions of operons were found to be fully transcribed after initiation (Nieß et al., 2017), other scenarios coincided, too. For instance, only subsets of operons may be transcribed, or even opposing transcription reads in a single operon occurred (Mao et al., 2015). Accordingly, we assumed that only 10% of the initiated operons are transcribed completely. In other words, 10% of experimentally observed initiated operons were anticipated to finish full operon transcription even outside of PFR. The majority (90%) of other gene transcriptions was assumed to stop immediately after RNAPs have reached individual gene ends. For the sake of comparability, only relative mRNA enrichments are depicted in Figure 7c, referring to the mRNA level of individual cells after they have fluctuated through the bioreactor for 180 s. This time point was chosen to visualize the spatial distribution of already adapted cells and the ones which are still influenced by regime changes (Figure 7).

2.4.3 | Translation

The translational process is modeled by describing the movement of ribosomes (RIB) on the mRNA strand. The process is assumed to take place in cotranscriptional manner (Proshkin et al., 2010): After

synthesized mRNA reached a minimum length of 80 nts (Δx_{RIB}) (Bremer & Dennis, 2008), the first ribosome attaches to the free 5'-cap end. Further ribosomes may bind too, provided that minimum distance between two subsequent ribosomes and the maximum number of translations per mRNA (max_{TL}) are fulfilled. Released ribosomes may be reused following the same scenario. The initiation trigger $TL_{i,j}$ for a translation j on mRNA i can be described as

$$TL_{i,j,on|off}(t) = \begin{cases} 1 & \text{if position of } RIB_{i,j-1} \text{is } >80 \text{ and } j < max_{TL}, \\ 0 & \text{else.} \end{cases} \tag{7}$$

The movement of ribosomes is analogous to the movement of RNAP as

$$\frac{dx_{RIB}}{dt} = \begin{cases} v_{RIB} & \text{if initiated,} \\ 0 & \text{else.} \end{cases} \tag{8}$$

With the translational elongation rate $v_{RIB} = v_{RNAP}$ (Nieß et al., 2017; Proshkin et al., 2010) and the number of maximum translations $max_{TL} = 11$ (Bremer & Dennis, 2008). Because one cluster of genes revealed rapid degradation after 50 s in PFR at 28 h, only one translation per transcript was assumed for this group of genes. As soon as the ribosomes pass the last nucleotide, the protein is released and is assigned to the group of bulk proteins.

2.4.4 | Degradation

Degradation of transcripts is initiated as soon as ribosomal protection of the 5'-cap vanishes, which is a co-translational process. The velocity of the RNASE was adapted to the gene length $L_{mRNA,i}$, that is, individual $v_{RNASE,i}$ was estimated considering the experimentally observed mRNA median lifetime $t_{deg,med}$ of 2.8 min in nutrient-rich and of 4.6 min in starvation zones (Chen et al., 2015). Degradation was initiated for mRNA i as

$$Deg_{i,j,on|off}(t) = \begin{cases} 1 & \text{if position of } RIB_{i,TL_{max}} \text{is } > 80, \\ 0 & \text{else.} \end{cases} \tag{9}$$

with the movement of RNASE.

$$\frac{dx_{RNASE}}{dt} = \begin{cases} v_{RNASE} & \text{if initiated,} \\ 0 & \text{else.} \end{cases} \tag{10}$$

$$v_{RNASE,i} = \frac{L_{mRNA,i}}{t_{deg,med} - \frac{TL_{max} * \Delta x_{RIB}}{v_{RIB}}} \tag{11}$$

For transcripts longer than ≈ 1000 nts, degradation is initiated already when transcription is not finished yet. Chen et al. (2015) found that 88 of 263 mRNAs showed lifetimes of the 5'-cap shorter than the synthesis time of the transcript. Accordingly, co-transcriptional degradation was considered for long transcripts.

Protein degradation is described using a constant rate degradation rate k_{deg} for the bulk proteins (Maurizi, 1992). First-order degradation kinetics were assumed depending on the nutrient condition, as

$$k_{deg} = \begin{cases} 0.01 h^{-1} \text{innutrientrichzones,} \\ 0.08 h^{-1} \text{instarvationzones.} \end{cases} \tag{12}$$

Consequently, bulk protein of a subsequent time step $t + 1$ equals

$$protein(t + 1) = protein(t) \times (1 - k_{deg}). \tag{13}$$

2.5 | Geometry and reactor setup

To consider a relevant industrial fed batch fermentation scenario, a 54-m^3 stirred tank bioreactor was chosen. The main geometry was derived from Haringa et al. (2016) with precise dimensions and information about the inner geometry from Kuschel et al. (2017). The reactor setup included four baffles and a stirrer with two Rushton stirrers equipped with eight blades at the bottom and six blades at the top. With a stirring rate of 100 rpm the tip speed of 6.75 m s^{-1} was reached. The impeller Reynolds number was 2.77×10^6 and the required power was 225.69 kW, equaling a power number of 13.64. The feeding rate was set to 3.68 kg m^{-3} s^{-1} for an average growth rate of 0.2 h^{-1}. Aeration, gas transfer, and oxygen uptake were neglected in the study. The simplifying focus on the mono-phase conditions mirrors the basic strategy to showcase the propagation of transcript dynamics and the occurrence of additional large-scale ATP demands. Noteworthy, experimental data of Löffler et al. (2016) were measured particularly excluding any impacts of oxygen limitation. Furthermore, said power inputs and cultivation conditions were chosen such that oxygen limitation is unlikely in the large-scale scenario. Cell concentration of 31.8 $kg_{CDW} \cdot m^{-3}$ was assumed and a simple Monod-like kinetic was used to simulate the substrate uptake q_s:

$$q_S = q_{S,max} \cdot \frac{c_S}{K_S + c_S}, \tag{14}$$

where $q_{s,max}$ is the maximum uptake rate, c_s is the glucose concentration, and the approximated substrate-specific uptake constant K_s with 4 mg L^{-1}. The maximum uptake rate was calculated with the biomass substrate yield $Y_{XS} = 0.25$ $g_s \cdot g^{-1}_{CDW}$ and the maximum growth rate $\mu = 0.2$ h^{-1} (Villadsen et al., 2011). Based on the experimental observations in Löffler et al. (2016) we concluded that stringent response is the predominant regulatory scheme initiated by repeated starvation. As a key characteristic stringent response reduces ATP consuming procedures trying to keep carbon supply on the maximum level achievable under stress conditions. Accordingly, we consider glucose uptake as a Monod-type function not being affected by the stringent response observed.

2.6 | Simulation setup

For the numerical simulation, the commercial calculation tool ANSYS Fluent version 19.1 was used. With 872,232 hexahedral numerical

cells resulting in an aspect ratio of 12.6 and a minimal orthogonal quality of 0.34, high mesh quality was achieved. Schmidt (*Sc*) number tuning with *Sc* = 0.2 leads to the same circulation time as achieved by Haringa et al. (2016). The flow field was approximated by solving the Reynolds-averaged Navier-Stokes (RANS) equations in combination with the realizable k-ε model for turbulence. All surfaces were set as no-slip boundaries except for the no-shear top area, which equaled the reactor filling height. Baffles and impellers were modeled as 0-thickness walls. Both impeller units were set to sliding mesh motion to generate a more realistic flow field. For glucose feed, a separate volume at the top of the reactor was defined, and a constant mass flow was set. The flow field and uptake kinetics were calculated every 10 ms until the glucose concentration was constant and a pseudo stationary gradient was reached, showing constant metabolic activity. The conditions were "frozen" for 180 s to track bacterial movements. These *lifelines* were simulated as massless Lagrangian particles with a discrete random walk (DRW) model passing through the flow field. Every 30 ms, the position and glucose concentration of each bacterium were recorded. In total, 120,000 bacterial cells were tracked over around 180 s (residence time distribution: Supporting Information Appendix Figure E14). According to the ergodic theorem, the same average values are obtained by tracking 1,080,000 bacteria for 20 s (for more information, see Appendix: Supporting information F). However, due to the limitation of simulation time and capacity, the simulation results were extended by repeating the single lifelines (Figure 2b) every 180 s while preserving the lifeline cluster groups. RNAP activities, mRNA levels, ribosomal activities, and protein formation were calculated as described in Section 2.4 considering each lifeline cluster (RNAP and mRNA profile of one lifeline group: Figure 2c; mRNA content of individual cells after 180 s: Figure 7c). Finally, additional ATP demands were estimated (Figure 2d).

3 | RESULTS

3.1 | Simulation results biological model

To identify data-driven parameters for the model of *E. coli* K12 W3110, clusters of mean mRNA levels were identified (Supporting information Appendix Section C). Figures 3–5 show the simulations (blue lines) and mean experimental values (red dots) for mRNA levels, active RNAPs, and the number of translated proteins per cell during PFR passage. The

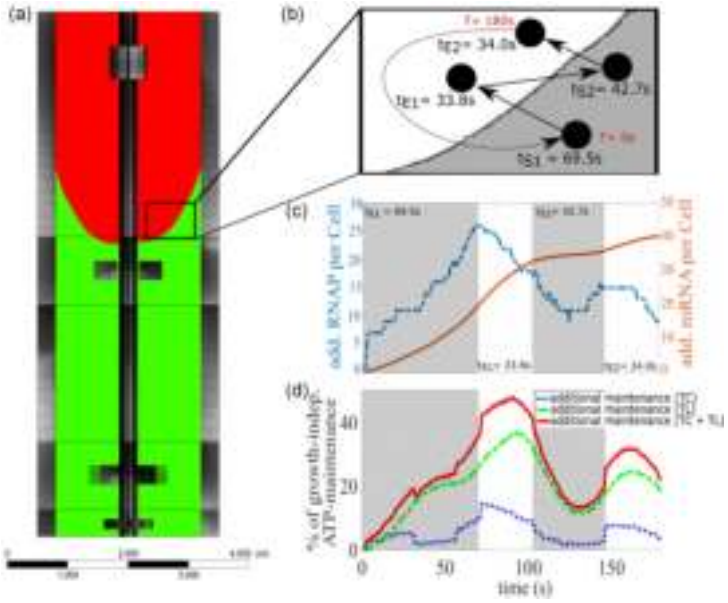


FIGURE 2 Impact of frequent exposure to feast and famine conditions in a large-scale bioreactor (a). White areas reflect nutrient excess, while gray areas indicate starvation. The size of the areas reflects the corresponding residence time indicated with t_S for starvation and t_E for excess residence time (c: Bar plot for one cluster of particle trajectories). The starvation-induced regulatory responses are propagated into the glucose excess zone, causing a maximum growth-independent ATP-maintenance in the glucose excess regime (d) based on additional active RNA polymerase(RNAP) for transcription (TC) (c) and ribosomes for translation (TL) [Color figure can be viewed at wileyonlinelibrary.com]

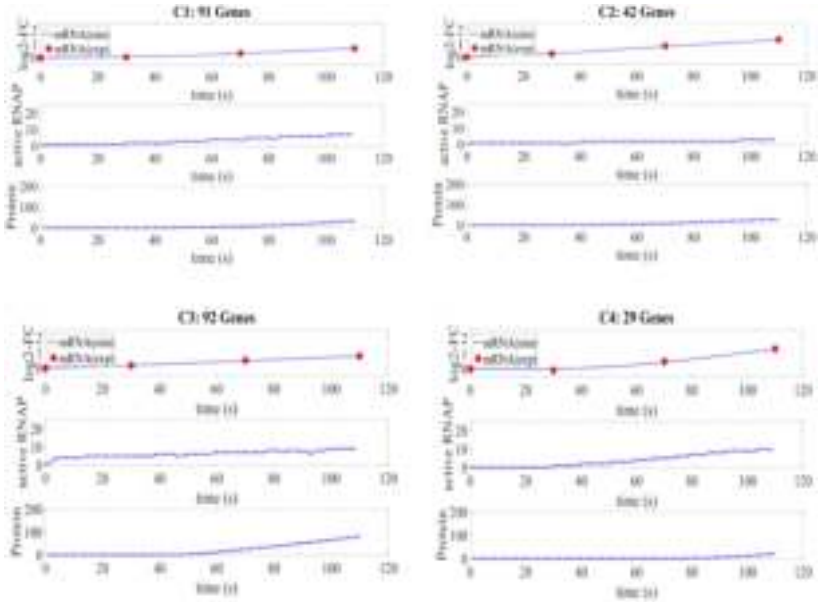


FIGURE 3 Simulated, cell-specific number of *additional* messenger RNA (mRNA) levels (red dots: experiments; blue line: simulation), active RNA polymerase (RNAP) and translated proteins of clusters 1–4 (C1–C4) along starvation passage (t = 0–110 s, Figure 1) at 25 min. The logarithmic fold change (log2-FC) relates to stirred tank reactor values at the same process time [Color figure can be viewed at wileyonlinelibrary.com]

synthesis rate accelerates over time as more RNAP molecules are involved in the transcriptional response to the starvation stimulus. After a transcript is completely synthesized and the RNAP is released the number of active RNAP shortly drops before it rises again. Synthesized proteins appear with a delay that corresponds to the required translation time. Activities for transcription and translation result in additional ATP demands which are indicated as add-ons to nongrowth-dependent maintenance (NAM), shown in Supporting information Appendix Figure B9. For transcription, costs are derived from nucleotide balancing, including the release of nucleotides by mRNA degradation and the need for mRNA synthesis. Cost for translation mirrors the amino acid needs and integration according to ribosomal activity. As indicated, translation costs are more than 2.5-fold higher than those of transcription. At maximum, cells have to bear 36.8% additional NAM, 10.4% coding for transcription, 26.4% for translation. This happens during the early phase of frequent starvation exposure, that is, after three starvation passages (25 min process time). After 2 h process time, the ATP demand still increases. More than 45% NAM increase is observed, illustrating the remaining high number of active RNAP. Later, after 28 h, NAM add-ons reduce more than fivefold compared with maximum needs. Then, transcription accounts for about 1% NAM rise only. The total NAM increase only mirrors 9.5%.

3.2 | Linking cluster kinetics

Cellular adaptations to frequent environmental stimuli are mirrored in the cluster dynamics of differentially expressed genes (DEGs) that were measured after 25 min, 2 h, and 28 h. Only 81 of 521 DEGs are conserved over the entire process time. This reflects the replacement of the initial sigma factor 70 dominated response by σ_{38} mediated regulatory programs (Löffler et al., 2016). To simulate the transition, so-called "damping" and "amplification" factors were identified using mean gene expressions as a reference based on the simulated log2FC mRNA dynamics in Figures 3–5. Clusters were subdivided in "persisting," "subsiding," and "non-active" fractions. The first collected genes with continuing high expression levels, while the second comprised genes with declining expression levels. The last summed those genes that were either not yet or no more expressed between subsequent time points (Figure 6). The damping factor is the ratio of the mean $\log_2$FC of subsiding and persisting genes for each cluster between two time points. The amplification factor is the ratio of the mean $\log_2$FC of genes at 25 min, which are active at time point 2 h divided by the mean $\log_2$FC ratio of genes activated after 2 h (see exemplary calculation Figure 6). The factors are used to calculate the amount of active RNAP:

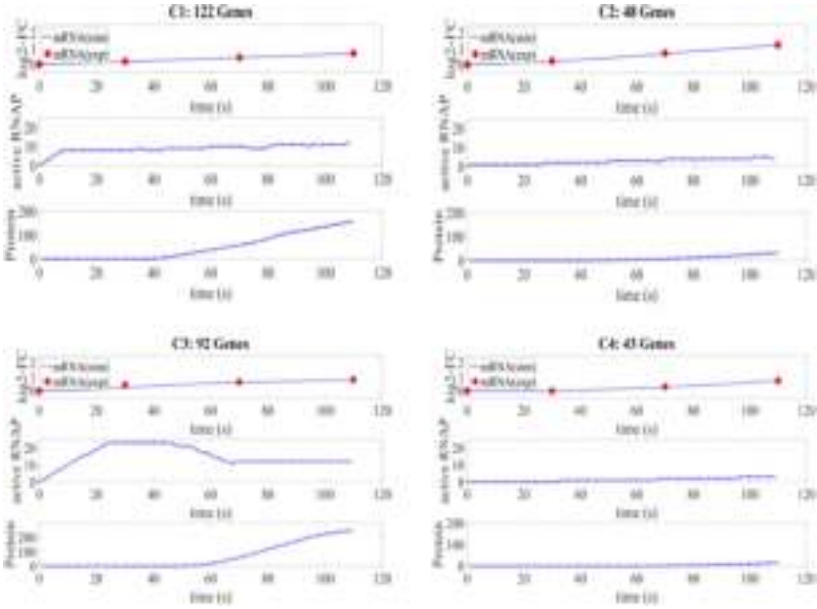


FIGURE 4 Simulated, cell-specific number of *additional* messenger RNA (mRNA) levels (red dots: experiments; blue line: simulation), active RNA polymerase (RNAP) and translated proteins of clusters 1–4 (C1–C4) along starvation passage (t = 0–110 s, Figure 1) at 2 h. The logarithmic fold change (log2-FC) relates to stirred tank reactor values at the same process time [Color figure can be viewed at wileyonlinelibrary.com]

$$x_{RNAP}(s,t) = x_{RNAP}(t)*\left(1 - \frac{Damping\ factor}{(S_{crit} - S_{min})}*(S(t) - S_{min})\right) \quad (15)$$

$$x_{RNAP}(s,t) = x_{RNAP}(t)*\left(\frac{Amplification\ factor}{(S_{crit} - S_{min})}*(S(t) - S_{min})\right) \quad (16)$$

The entire transition process is guided by the number of starvation passages $S(t)$ per time. S_{min} encodes the minimum and S_{crit} the critical number of passages. Whereas the first is a regression parameter, the later reflects experimental observations of Löffler et al. (2016) as follows: 25 min equal 3 PFR (starvation) passages, 2 h equal 14, and 28 h equal 176. Noteworthy, the 28 h benchmark is chosen as a new steady state was observed already then (Löffler et al., 2016). The modeling approach allows transferring of the STR/PFR observations to other conditions using the frequency of feast/famine exposure $S(t)$ as a key criterion.

3.3 | Numerical simulation

3.3.1 | Glucose gradient

Applying the criterion of converged turbulent dissipation rate/power input, the pseudo-stationary glucose gradient of Figure 7 was obtained (Figure 7). Accordingly, no further changes in glucose concentrations simulated at five locations occurred. The average concentration in the bioreactor was 23.74 mg L^{-1}. For comparison, the average glucose level observed by the Lagrangian particles ("cells") was 22.79 mg L^{-1}. Consequently, only 4% deviation was found, which is qualified as a small difference indicating good homogeneous distribution and reflecting impacts of the turbulence model and of particle lifeline filtering. The volumetric distribution between starvation and excess zone is 73%–27%, respectively. Again, similar percentages were calculated by integrating mean residence times of all lifelines. The mean residence time of the cells in the starvation regime is 9.46 s (Supporting Information Appendix Figure E14), which is in the same range as published by Haringa et al. (2016).

Ideally, large-scale simulations should have been compared with real in situ measurements to challenge the predictions. However, such data are missing, which represents a common problem often faced by academic groups. Nevertheless, applying CFD simulations still offers the best chances for getting highly accurate large-scale predictions as complex hydrodynamics, even including overlapping flow fields between stirrers, are well predictable. Notably, the latter may hamper the application of simplifying compartment-based estimations, which basically assume separated flow fields between stirrers (see Supporting Information Appendix H).

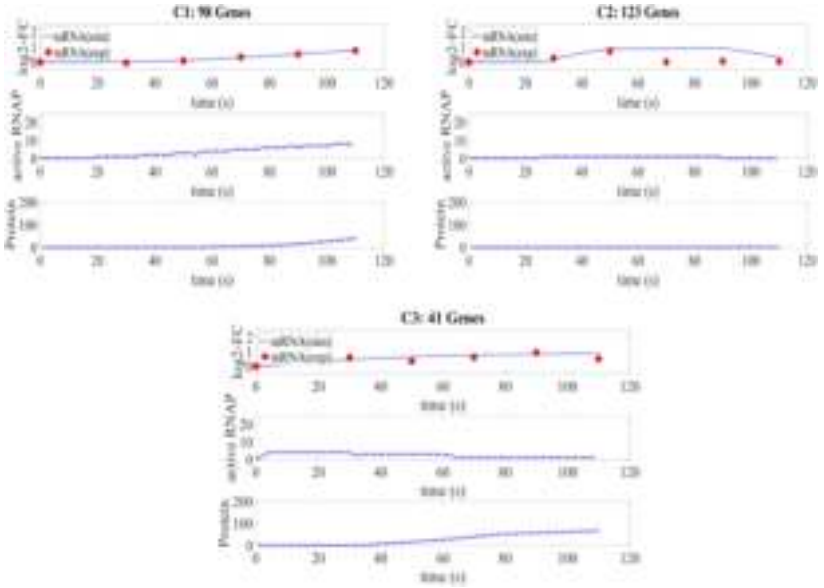


FIGURE 5 Simulated, cell-specific number of additional messenger RNA (mRNA) levels (red dots: experiments; blue line: simulation), active RNA polymerase (RNAP) and translated proteins of clusters 1–3 (C1–C3) along starvation passage (t = 0–110 s, Figure 1) at 28 h. The logarithmic fold change (log2-FC) relates to stirred tank reactor values at the same process time [Color figure can be viewed at wileyonlinelibrary.com]

For simplicity, Eulerian simulations only considered the liquid phase, thereby assuming sufficient oxygen supply in the bioreactor without calculating DO concentrations explicitly. Furthermore, additional turbulence due to bubble interaction was neglected, too. Substrate consumption followed Monod-type kinetics taking place in each numerical cell. This implied that bacterial cells were homogeneously distributed at each time step.

3.3.2 | Statistical evaluation

Lifeline statistics reflect the imprint of changing micro-environmental conditions on cells fluctuating through the bioreactor. To be precise, cellular residence times in different concentration zones and shifts between proximal regimes were studied. At the start, cells were "inserted" into the bioreactor along a straight line reaching from top to bottom. After a few simulation steps, cells were distributed homogeneously before individual cell tracking started for 180 s. Lifeline records were cured by percolating only those with residence times longer than 0.13 s. The latter represent unrealistically turbulent fluctuations. The following threshold was defined for regime analysis: If cells experience lower or higher glucose levels than K_S for at least one second, the period is labeled as starvation or saturation time, respectively. Noteworthy, the minimum residence time of 1 s equals the average metabolite turnover time in *E. coli* (Shamir et al., 2016; Taymaz-Nikerel et al.,2011). The implementation of the harsh regime boundary K_S finally leads to rapid and somewhat artificial regime shifts. They were excluded from analysis by ignoring the upper and lower 1% of regime changes. Alternately, the consideration of alarmones such as (p)ppGpp serving as intracellular triggers to initiate transcriptional regulation may yield continuous models. Unfortunately, understanding of alarmone formulation, degradation, and alarmone induced regulation is still too fragmented to build dynamic transcriptional models. In total, measures for residence time percolation and shift filtering only reduced the data set by 3% (Supporting Information Appendix Figure E14).

At maximum, 41 regime shifts were observed during the 180 seconds observation period. Most frequently, 20 regime changes occurred and cells rested in a single zone no longer than 30 seconds. As a key characteristic, cells once exposed to glucose starvation reset their regulation signal. But RNAP and ribosomes remain active, propagating the starvation response into the glucose excess regime (Figure 2c,d). According to their starvation pattern the cell lifelines were assigned to 70 different clusters. Thereby each cluster represents a specific fluctuation pattern, reflecting the amount and duration of changes between starvation and excess zone (Figure 2b).

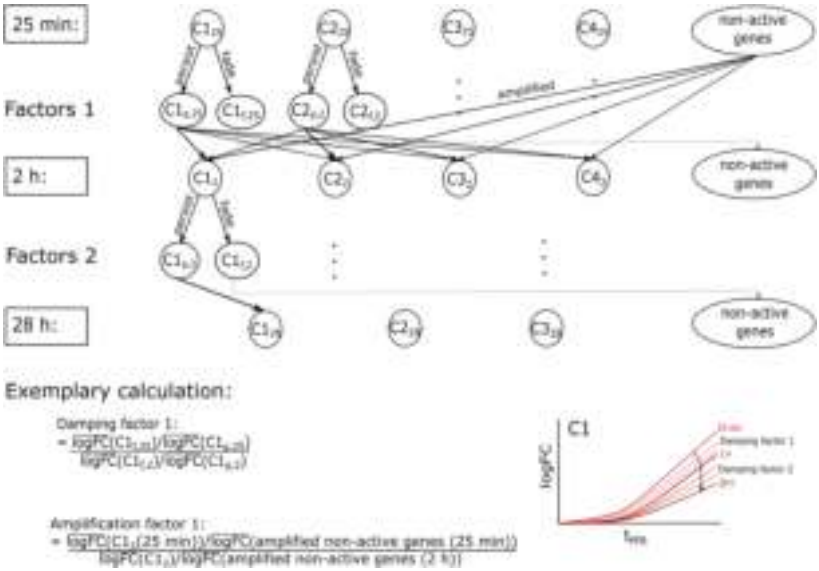


FIGURE 6 Scheme illustrating how "damping" and "amplification" factors are derived from experimental cluster data (C1–C4) measured at 25 min, 2 h, and 28 h after plug-flow reactor passage. In general, active genes of individual clusters may continue amplification (persisting, *p*), reduce expression (fading, *f*), or even be activated from the group of non-active genes. To bridge the gene expression dynamics from 25 min to 2 h and from 2 to 28 h, so-called damping and amplification factors are calculated as illustrated in the example. They use mean logFC values of the relevant time points. The damping factor correlates the ratios of fading-to-persisting genes of the two subsequent measurements. By analogy, the amplification factor calculates ratios of mean numbers of gene expression versus "first time" amplified genes for each time point and correlates the same for two subsequent events [Color figure can be viewed at wileyonlinelibrary.com]

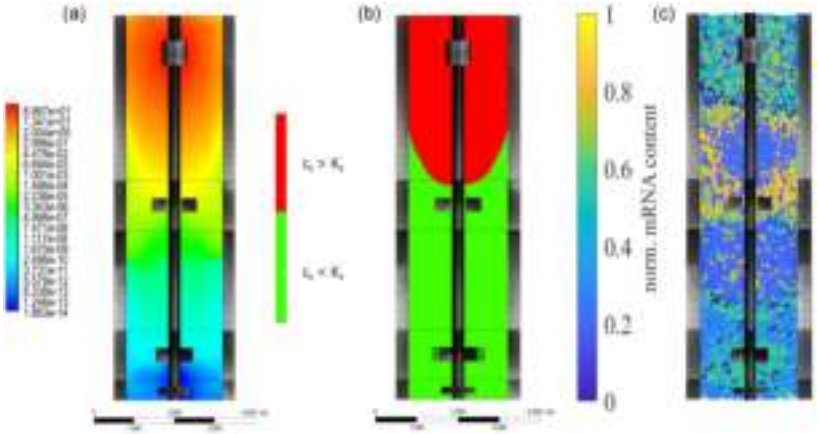


FIGURE 7 Simulation results of the 54 m^3 stirred tank reactor. (a, left) Log-contours of c_s/K_s gradients. (b, middle) Assignment to regimes (limitation, green: $c_s < K_s$; saturation, red: $c_s > K_s$). (c, right) Normalized messenger RNA (mRNA) content analyzed after cells fluctuated 180 s through the bioreactor [Color figure can be viewed at wileyonlinelibrary.com]

3.3.3 | Coupling the biological model with lifelines

To minimize the computational efforts, particle lifelines were exported from ANSYS Fluent. Starvation patterns of lifelines (Figure 2b) served as input for the biological model. A workflow scheme is provided in the Supporting Information Appendix Figure H16. It was assumed that each entry of the starvation zone activated the expression of distinct gene clusters as experimentally observed. Accordingly, individual expression patterns were estimated for each cell, mirroring their particular tracking history.

As indicated in Figure 7c, the basic expression level of the cellular population is increased by 37.7% compared with the reference. This reflects the additional cellular needs to adapt to the heterogeneous mixing conditions in the reactor. Noteworthy, high mRNA levels, induced by preceding starvation, are propagated into glucose-rich zones (Figure 7c). Expression patterns reflecting starving conditions occur in saturating glucose zones and vice versa. The phenomenon mirrors the delayed transcriptional response that is slower than convective movements of cells in the bioreactor. Consequently, a high transcriptional heterogeneity occurred in the tank. A maybe surprising pattern is disclosed: In the lower part of the reactor, cells envisage low glucose concentrations but show reduced mRNA levels (norm. mean mRNA level: 0.33). On the contrary, cells facing high glucose levels in the upper part reveal high mRNA levels (norm. mean mRNA level: 0.42).

About 25% of the cells permanently stay in the starvation zone. This fraction even adapts to the limiting conditions, which reduce the transcriptional response gradually. Cells located close to zones of glucose excess highly fluctuate between starving and saturating conditions. Consequently, strong gene expression responses are observed.

The average ATP-demand of a newborn, not preconditioned population of 120,000 cells exposed to the 54 m^3 bioreactor is shown in Figure 8. Basically, plot 8 illustrates the cyclic passing of 180 s lasting lifelines. To filter related peaks, only average values are indicated using a moving median filter over 700 data points. Furthermore, a moving standard deviation with a window size of 400 data points is added as a shadow. Synchronization-like patterns reflect the clustering of particles in groups. 70 bins were used with passable computational effort. The maximum of 45% additional maintenance is predicted shortly (0.03 h) after cells were exposed to the bioreactor condition. After about 0.5 h most of the population has adapted to the heterogeneous environment reducing the additional ATP demand to 6.5%. After 4.22 h the last cell fraction is in the adaption state.

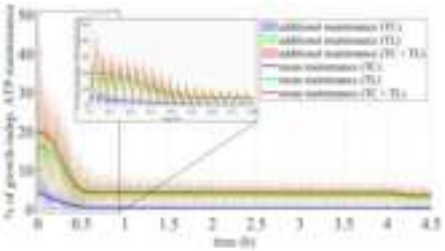


FIGURE 8 Additional ATP demands of a population of 120,000 "newborn," not preconditioned cells in a 54 m^3 reactor over 4.5 h process time. Courses for mean transcription (blue: TC), translation, (green: TL) and the sum of both are depicted (red: TC + TL). The shaded areas display the standard deviation [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Parameters and omitted genes of the SQBC algorithm

Time point	25 min	2 h	28 h
ntry	1000	1000	1000
rmax	0.5	0.5	0.9
Min size	3	3	5
Genes omitted	0	1	6

Abbreviations: *Ntry*, number of trials per iteration; *rmax*, maximum radius as a proxy for correlation; *Min size*, minimum number of observations per clusters.

4 | DISCUSSION

To disclose spatial heterogeneities of regulation patterns and additional ATP demands of *E. coli* K12 W3110 exposed to a 54 m^3 STR CFD based lifeline analysis was coupled with agent-based modeling for transcription and translation.

As a prerequisite, a proper model describing stress-induced dynamics of transcription and translation is needed. Applying a clustering approach, it was possible to properly describe the experimentally observed transcription dynamics (Löffler et al., 2016) of 821 genes using 16 parameters. Under steady-state conditions, newly synthesized and recycled *nts* equilibrate in cells before they enter PFR. However, ATP demands transcription rise inside PFR as mRNA synthesis exceeds the recycling rate. The introduction of 'amplification' and 'damping' factors managed to model the transition from fast response to long-term adaptation, as visible in the experimental data. Accordingly, modeling succeeded to mirror the cellular

TABLE 2 ATP costs for de novo nucleotide synthesis, based on Löffler et al. (2016)

Base	ATP	NADH	NADPH	Overall ATP
Guanin	11	−3	1	8.53
Cytosin	13	−3	0	9
Adenin	9	0	1	6.53
Uracil	7	0	1	7
Average	10	−1.5	0.75	7.7

efforts shifting control from σ_{70} to σ_{38}, more and more (Löffler et al., 2016). Given that *E. coli* cells with doubling times of 3.3 h contain about 120 active RNAPs (Bremer & Dennis, 2008), approximately one-fifth of the available RNAPs at 25 min is used in the transcriptional response (Figure 3).

The number of involved RNAPs slowed down after 28 h for two reasons. First, the absolute number of initiated transcripts is reduced, which mirrors cellular adaptation. Second, transcription even stopped after 50 s for a large group of genes (Figure 5). Accordingly, reduced synthesis costs occur. This is amplified by the prolonged lifetime of transcripts during starvation, which further reduces the amount of synthesis to obtain a certain level of mRNA abundance as described in Section 2.4.4 mRNA lifetime is proportional to their distance from the 5′ end of the transcript according to Chen et al. (2015), which is in line with the observation that the 5′ end contains important determinants that regulate RNA lifetime (Arnold et al., 1998).

Supporting Information Appendix Figure B9 shows that protein synthesis accounts for the major part of ATP consumption. Peak values reach about 45% NAM at 2 h process time. Later on, the demand steadily reduces as less mRNA transcripts are synthesized and less proteins are translated. At 25 min, the number of active ribosomes involved in the stress response rises steadily during starvation, reaching 10% of all available ribosomes (Reference: 10^3 ribosomes/cell; Bremer & Dennis, 2008). The fraction decreases to 5% after 28 h reflecting the adaptation process.

Linking the transcription and translation model to large-scale lifelines reveals the impact of delayed cellular response to fast external changes. In essence, cellular responses of transcription and translation are slower than convective zone shifts. Consequently, they are transported from one location to another, basically decoupled from external changes. Spatial analysis of all cells after 180 s (Figure 7c) reveals the highest transcript levels close to or even inside the glucose excess regime (50%–80% mRNA upregulation), while the lowest are found at the bottom. Once initiated, the starvation response propagated into the glucose excess zone. There, additional needs for transcription, translation, and ATP may limit the targeted formation of industrial products in microbial cells.

Noteworthy, it is exactly this feature that distinguishes the current model from previous lifeline studies (Haringa et al., 2016, 2017; Kuschel et al., 2017). There, metabolic and cell cycle responses were considered as an *instantaneous* cellular responses. Fast external changes are immediately translated into cellular replies. Later, Haringa et al. (2018) implemented metabolically buffered responses by considering variable enzyme pools (Tang et al., 2017). In this context, the current study proceeds by additionally integrating downstream transcriptional responses incorporating another level of cellular control. Our approach introduces the *non-instantaneous*, delayed response by considering intracellular programs of longer time scales than external changes. Accordingly, responses may propagate in different zones of the reactor, causing nonexpected transcriptional regulation programs there.

The approach was exploited further by estimating the entire add-on ATP demand for 120,000 newborn cells monitoring 4.5 h process time (Figure 8). As shown, the adaptation of the population is finished after 4.2 h disclosing a remaining ATP add-on of about 6% NAM compared to the 45% max NAM at the beginning. In terms of microbial productivity, these ATP needs simply reduce the amount of available ATP for product formation; that is, they limit biomass-specific productivities. The phenomenon has often be described in large-scale fermentation (Lara et al., 2006). Noteworthy, it is likely to be pronounced in hyper-producing cells with ATP intensive product formation. Often enough, such production processes run in fed-batch mode, installing reduced, limiting metabolic activity to stay within the technical limits of aeration and cooling. Consequently, those additional ATP needs hit cells with limited ATP forming capacities.

To evaluate the impact of particle simulation time, an additional simulation was conducted using a high-performance computation cluster studying 60,000 particles for about 460 s. Similar results were obtained for the key readouts, that is, time courses of ATP maintenance demands and residence time distributions remained. The simulated adaptation time reduced from 4.2 to 3.7 h, mirroring the lowered amount of particles staying in the starvation zone for the entire process time (around 5%) (see Supporting Information J). However, modelers need to consider that long simulation times are likely to violate the intrinsic constraint of one-way coupling, neglecting particle-environment interactions for the sake of simplicity. In this dilemma, we decided for the analysis of 180 s to capture key dynamics while still fulfilling the one-way coupling constraint.

As pointed out by Löffler et al. (2016), the majority of transcription dynamics are caused by the frequent on/off switching of stringent response, mediated by rising intracellular (p)ppGpp levels. Hence, creating stringent response deficient strains (Michalowski et al., 2017) opens the door to prevent non-wanted NAM increase. Besides, other cellular stress programs may be targeted as well (Supporting Information Appendix Table D4).

5 | CONCLUSION

The current modeling approach marries computational lifeline analysis with cellular regulation models, thereby introducing the *non-instantaneous* cellular response to changing extracellular conditions. Consequently, the spot of stress induction and the location of cellular phenotype do not need to be the same. Accordingly, heterogeneities in large-scale bioreactors comprise the physical levels linking local conditions tightly with metabolic responses and the cellular regulation level encompassing delayed responses such as transcriptional or translational effects.

To detect the latter and to describe them properly in data-driven models, experimental scale-up simulators are necessary that mirror transcriptional and translational cellular replies as performed by Kuschel and Takors (2020). The setting of such devices may differ from “conventional” scale-up simulators that typically mimic circulation times. Because the entire transcriptional responses should be clearly detectable, rather long stress exposure periods should be installed and read.

ACKNOWLEDGMENTS

The authors would like to thank the group of Computational Biology at the Institute of Biochemical Engineering for the use of the Galaxy-Server and Maike Kuschel and Flora Siebler for their support and inspiring discussions during the simulation process. Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Prof. Dr.-Ing. Ralf Takors advised the study, while Moritz Wild implemented major parts of the agent-based model. Simulation, evaluation, and writing of the manuscript were conducted by Julia Zieringer.

ORCID

Ralf Takors http://orcid.org/0000-0001-5837-6906

REFERENCES

Arnold, T. E., Yu, J., & Belasco, J. G. (1998). mRNA stabilization by the ompa 5' untranslated region: Two protective elements hinder distinct pathways for mRNA degradation. *RNA*, *4*(3), 319–330.

Bremer, H., & Dennis, P. P. (2008). Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, *3*(1), 1–49. https://doi.org/10.1128/ecosal.5.2.3

Chen, H., Shiroguchi, K., Ge, H., & Xie, X. S. (2015). Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Molecular Systems Biology*, *11*(1), 781. https://doi.org/10.15252/msb.20145794

Delvigne, F., Takors, R., Mudde, R., van Gulik, W., & Noorman, H. (2017). Bioprocess scale-up/down as integrative enabling technology: From fluid mechanics to systems biology and beyond. *Microbial Biotechnology*, *10*(5), 1267–1274. https://doi.org/10.1111/1751-7915.12803

Garcia-Ochoa F., & Gomez E. 2009. Bioreactor-scale-up and oxygen transfer in microbial processes: An overview. *Biotechnology Advances*, *27*, 153–176. https://doi.org/10.1016/j.biotechadv.2008.10.006

Haringa, C., Deshmukh, A. T., Mudde, R. F., & Noorman, H. J. (2017). Euler-Lagrange analysis towards representative down-scaling of a 22 m3 aerobic S. cerevisiae fermentation. *Chemical Engineering Science*, *16*, 652–663. https://doi.org/10.1016/j.ces.2017.01.014

Haringa, C., Tang, W., Deshmukh, A. T., Xia, J., Reuss, M., Heijnen, J. J., Mudde, R. F., & Noorman, H. J. (2016). Euler-Lagrange computational fluid dynamics for (bio) reactor scale down: An analysis of organism lifelines. *Engineering in Life Sciences*, *16*(7), 652–663. https://doi.org/10.1002/elsc.201600061

Haringa, C., Tang, W., Wang, G., Deshmukh, A. T., van Winden, W. A., Chu, J., van Gulik, W. M., Heijnen, J. J., Mudde, R. F., & Noorman, H. J. (2018). Computational fluid dynamics simulation of an industrial *P. chrysogenum* fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. *Chemical Engineering Science*, *175*, 12–24. https://doi.org/10.1016/j.ces.2017.09.020

Kaleta, C., Schäuble, S., Rinas, U., & Schuster, S. (2013). Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnology Journal*, *8*(9), 1105–1114. https://doi.org/10.1002/biot.201200267

Kibbe, W. A. (2007). Oligocalc: an online oligonucleotide properties calculator. *Nucleic Acids Research*, *35*(Web Server issue), W43–W46. https://doi.org/10.1093/nar/gkm234

Kuschel, M., Siebler, F., & Takors, R. (2017). Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. *Bioengineering*, *4*(2), 27. https://doi.org/10.3390/bioengineering402002

Kuschel, M., & Takors, R. (2020). Simulated oxygen and glucose gradients as a prerequisite for predicting industrial scale performance a priori. *Biotechnology and Bioengineering*, *117*, 2760–2770. https://doi.org/10.1002/bit.27457

Lapin, A., Schmid, J., & Reuss, M. (2006). Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor: A structured–segregated approach. *Chemical Engineering Science*, *61*(14), 4783–4797. https://doi.org/10.1016/j.ces.2006.03.003

Lara, A. R., Galindo, E., Ramírez, O. T., & Palomares, L. A. (2006). Living with heterogeneities in bioreactors: Understanding the effects of environmental gradients on cells. *Molecular Biotechnology*, *34*(3), 355–382. https://doi.org/10.1385/MB:34:3:355

Leisch, F. (2006). A toolbox for k-centroids cluster analysis. *Computational Statistics and Data Analysis*, *51*(2), 526–544. https://doi.org/10.1016/j.csda.2005.10.006

Löffler, M., Simen, J. D., Jäger, G., Schäferhoff, K., Freund, A., & Takors, R. (2016). Engineering *E. coli* for large-scale production–strategies considering atp expenses and transcriptional responses. *Metabolic Engineering*, *38*, 73–85. https://doi.org/10.1016/j.ymben.2016.06.008

Mao, X., Ma, Q., Liu, B., Chen, X., Zhang, H., & Xu, Y. (2015). Revisiting operons: an analysis of the landscape of transcriptional units in *E. coli*. *BMC Bioinformatics*, *16*(1), 356. https://doi.org/10.1186/s12859-015-0805-8

Maurizi, M. R. (1992). Proteases and protein degradation in *Escherichia coli*. *Experientia*, *48*(2), 178–201. https://doi.org/10.1007/BF01923511

Michalowski, A., Siemann-Herzberg, M., & Takors, R. (2017). *Escherichia coli* HGT: Engineered for high glucose throughput even under slowly growing or resting conditions. *Metabolic Engineering*, *40*, 93–103. https://doi.org/10.1016/j.ymben.2017.01.005

Neidhardt, F. C., Ingraham, J. L., & Schaechter, M. (1990). *Physiology of the bacterial cell: A molecular approach*. Sinauer.

Neubauer, P., & Junne, S. (2010). Scale-down simulators for metabolic analysis of large-scale bioprocesses. *Current Opinion in Biotechnology*, *21*(1), 114–121. https://doi.org/10.1016/j.copbio.2010.02.001

Neubauer, P., & Junne, S. (2016). Scale-up and scale-down methodologies for bioreactors. In *Bioreactors: Design, operation and novel applications*. (pp. 323–354) Wiley. https://doi.org/10.1002/9783527683369.ch11

Nieß, A., Löffler, M., Simen, J. D., & Takors, R. (2017). Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of *E. coli* cultures in large-scale bioreactors: Experimental evidence and mathematical model. *Frontiers in Microbiology*, *8*, 1195. https://doi.org/10.3389/fmicb.2017.01195

Noorman, H. J. (2011). An industrial perspective on bioreactor scale-down: What we can learn from combined large-scale bioprocess and model fluid studies. *Biotechnology Journal*, *6*(8), 934–943. https://doi.org/10.1002/biot.201000406

Noorman, H. J., & Heijnen, J. J. (2017). Biochemical engineering's grand adventure. *Chemical Engineering Science*, *170*, 677–693. https://doi.org/10.1016/j.ces.2016.12.065

Oosterhuis, N. M. G., & Kossen, N. W. (1983). Dissolved oxygen concentration profiles in a production-scale bioreactor. *Biotechnology and Bioengineering*, *26*, 546–550. https://doi.org/10.1002/bit.260260522

Proshkin, S., Rahmouni, A. R., Mironov, A., & Nudler, E. (2010). Cooperation between translating ribosomes and rna polymerase in transcription elongation. *Science*, *328*(5977), 504–508. https://doi.org/10.1126/science.1184939

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edger: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Shamir, M., Bar-On, Y., Phillips, R., & Milo, R. (2016). Snapshot: Timescales in cell biology. *Cell*, *164*(6), 1302. https://doi.org/10.1016/j.cell.2016.02.05

Stouthamer, A. H. (1973). A theoretical study on the amount of atp required for synthesis of microbial cell material. *Antonie Van Leeuwenhoek*, *39*(1), 545–565. https://doi.org/10.1007/BF02578899

Takors, R. (2012). Scale-up of microbial processes: Impacts, tools and open questions. *Journal of Biotechnology*, *160*(1), 3–9. https://doi.org/10.1016/j.jbiotec.2011.12.010

Tang, W., Deshmukh, A. T., Haringa, C., Wang, G., van Gulik, W., van Winden, W., Reuss, M., Heijnen, J. J., Xia, J., Chu, J., Noormann, H. J. (2017). A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by *Penicillium chrysogenum*. *Biotechnology Bioengineering*, *114*(8), 1733–1743. https://doi.org/10.1002/bit.26294

Taymaz-Nikerel, H., Borujeni, A. E., Verheijen, P. J. T., Heijnen, J. J., & van Gulik, W. M. (2010). Genome-derived minimal metabolic models for *Escherichia coli* mg1655 with estimated in vivo respiratory atp stoichiometry. *Biotechnology and Bioengineering*, *107*(2), 369–381. https://doi.org/10.1002/bit.22802

Taymaz-Nikerel, H., Van Gulik, W. M., & Heijnen, J. J. (2011). *Escherichia coli* responds with a rapid and large change in growth rate upon a shift from glucose-limited to glucose-excess conditions. *Metabolic Engineering*, *13*(3), 307–318. https://doi.org/10.1016/j.ymben.2011.03.003

Villadsen, J., Nielsen, J., & Lidén, G. (2011). *Bioreaction engineering principles*. Springer Science & Business Media.

Zieringer, J., & Takors, R. (2018). In silico prediction of large-scale microbial production performance: Constraints for getting proper data-driven models. *Computational and Structural Biotechnology Journal*, *16*, 246–256. https://doi.org/10.1016/j.csbj.2018.06.002

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Zieringer J, Wild M, Takors R. Data-driven in silico prediction of regulation heterogeneity and ATP demands of *Escherichia coli* in large-scale bioreactors. *Biotechnology and Bioengineering*. 2020;1–14. https://doi.org/10.1002/bit.27568

Bibliography

Abbas-Aghababazadeh, F., Li, Q., and Fridley, B. L. (2018). "Comparison of normalization approaches for gene expression studies completed with high-throughput sequencing". In: *PloS one* 13.10.

Aghabozorgi, S., Seyed Shirkhorshidi, A., and Ying Wah, T. (2015). "Time-series clustering – A decade review". In: *Information Systems* 53, pp. 16–38.

Ahlstedt, H. and Lahtinen, M. (1996). "Calculation of flow field in a stirred tank with Rushton turbine impeller". In: *Proceedings of the Third CFX International Users Conference, Chesham, UK*. Vol. 30, pp. 91–108.

Aizenman, E., Engelberg-Kulka, H., and Glaser, G. (1996). "An Escherichia coli chromosomal" addiction module" regulated by guanosine [corrected] 3′, 5′-bispyrophosphate: a model for programmed bacterial cell death". In: *Proceedings of the National Academy of Sciences* 93.12, pp. 6059–6063.

Alibhai, M. and Villafranca, J. J. (1994). "Kinetic and mutagenic studies of the role of the active site residues Asp-50 and Glu-327 of Escherichia coli glutamine synthetase". In: *Biochemistry* 33.3, pp. 682–686.

Allison, K. R., Brynildsen, M. P., and Collins, J. J. (2011). "Heterogeneous bacterial persisters and engineering approaches to eliminate them". In: *Current opinion in microbiology* 14.5, pp. 593–598.

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., and Jirstrand, M. (2014). "Kinetic models in industrial biotechnology–improving cell factory performance". In: *Metabolic Engineering* 24, pp. 38–60.

Anders, S. and Huber, W. (2010). "Differential expression analysis for sequence count data". In: *Nature Precedings*, pp. 1–1.

Anders, S., Pyl, P. T., and Huber, W. (2015). "HTSeq—a Python framework to work with high-throughput sequencing data". In: *Bioinformatics* 31.2, pp. 166–169.

Andrews, S. C. and Guest, J. (1988). "Nucleotide sequence of the gene encoding the GMP reductase of Escherichia coli K12". In: *Biochemical Journal* 255.1, pp. 35–43.

Ankenbauer, A., Schäfer, R. A., Viegas, S. C., Pobre, V., Voß, B., Arraiano, C. M., and Takors, R. (2020). "Pseudomonas putida KT2440 is naturally endowed to withstand industrial-scale stress conditions". In: *Microbial biotechnology* 13.4, pp. 1145–1161.

Arnold, T. E., Yu, J., and Belasco, J. G. (1998). "mRNA stabilization by the ompA 5′untranslated region: two protective elements hinder distinct pathways for mRNA degradation." In: *Rna* 4.3, pp. 319–330.

Artsimovitch, I. and Landick, R. (2000). "Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals". In: *Proceedings of the National Academy of Sciences of the United States of America* 97.13, pp. 7090–7095.

Atherly, A. G. (1979). "Escherichia coli mutant containing a large deletion from relA to argA." In: *Journal of bacteriology* 138.2, pp. 530–534.

Atkinson, D. E. (1968). "Energy charge of the adenylate pool as a regulatory parameter. Interaction with feedback modifiers". In: *Biochemistry* 7.11, pp. 4030–4034.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006). "Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection". In: *Molecular systems biology* 2.1, pp. 2006–0008.

Backhaus, K., Erichson, B., Plinke, W., and Weiber, R. (2018). *Multivariate Analysemethoden*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M., Ramadan, H. A., Saini, K. S., Baeshen, N. A., and Redwan, E. M. (2015). "Production of biopharmaceuticals in E. coli: current scenario and future perspectives". In: *Journal of microbiology and biotechnology* 25.7, pp. 953–962.

Baeshen, N. A., Baeshen, M. N., Sheikh, A., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A., Saini, K. S., and Redwan, E. M. (2014). "Cell factories for insulin production". In: *Microbial cell factories* 13.1, p. 141.

Bailey, J. E. (1998). "Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities". In: *Biotechnology Progress* 14.1, pp. 8–20.

Bailey, J. (1991). "Toward a science of metabolic engineering". In: *Science* 252.5013, pp. 1668–1675.

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., and Leibler, S. (2004). "Bacterial persistence as a phenotypic switch". In: *Science* 305.5690, pp. 1622–1625.

Barker, C. S., Prüß, B. M., and Matsumura, P. (2004). "Increased motility of Escherichia coli by insertion sequence element integration into the regulatory region of the flhD operon". In: *Journal of bacteriology* 186.22, pp. 7529–7537.

Barker, M. M., Gaal, T., Josaitis, C. A., and Gourse, R. L. (2001). "Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro". In: *Journal of molecular biology* 305.4, pp. 673–688.

Barrios, A. F. G., Zuo, R., Hashimoto, Y., Yang, L., Bentley, W. E., and Wood, T. K. (2006). "Autoinducer 2 controls biofilm formation in Escherichia coli through a novel motility quorum-sensing regulator (MqsR, B3022)". In: *Journal of bacteriology* 188.1, pp. 305–316.

Basu, S., Yudkin, J. S., Kehlenbrink, S., Davies, J. I., Wild, S. H., Lipska, K. J., Sussman, J. B., and Beran, D. (2019). "Estimation of global insulin use for type 2 diabetes, 2018–30: a microsimulation analysis". In: *The Lancet Diabetes & Endocrinology* 7.1, pp. 25–33.

Becker, J., Zelder, O., Häfner, S., Schröder, H., and Wittmann, C. (2011). "From zero to hero. Design-based systems metabolic engineering of Corynebacterium glutamicum for L-lysine production". In: *Metabolic Engineering* 13.2, pp. 159–168.

Becker, P. J., Puel, F., Chevalier, Y., and Sheibat-Othman, N. (2014). "Monitoring silicone oil droplets during emulsification in stirred vessel: Effect of dispersed phase concentration and viscosity". In: *The Canadian Journal of Chemical Engineering* 92.2, pp. 296–306.

Bekker, M., De Vries, S., Ter Beek, A., Hellingwerf, K., and De Mattos, M. T. (2009). "Respiration of Escherichia coli can be fully uncoupled via the nonelectrogenic terminal cytochrome bd-II oxidase". In: *Journal of bacteriology* 191.17, pp. 5510–5517.

Belasco, J. G. and Higgins, C. F. (1988). "Mechanisms of mRNA decay in bacteria: a perspective". In: *Gene* 72, pp. 15–23.

Benjamini, Y. and Hochberg, Y. (1995). "Controlling the false discovery rate: a practical and powerful approach to multiple testing". In: *Journal of the Royal statistical society: series B (Methodological)* 57.1, pp. 289–300.

Berry, H. (2002). "Monte Carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation". In: *Biophysical Journal* 83.4, pp. 1891–1901.

Bezzo, F., Macchietto, S., and Pantelides, C. (2003). "General hybrid multizonal/CFD approach for bioreactor modeling". In: *AIChE Journal* 49.8, pp. 2133–2148.

Bi, E. and Lutkenhaus, J. (1993). "Cell division inhibitors SulA and MinCD prevent formation of the FtsZ ring." In: *Journal of bacteriology* 175.4, pp. 1118–1125.

Binder, D., Drepper, T., Jaeger, K.-E., Delvigne, F., Wiechert, W., Kohlheyer, D., and Grünberger, A. (2017). "Homogenizing bacterial cell factories: analysis and engineering of phenotypic heterogeneity". In: *Metabolic engineering* 42, pp. 145–156.

Boecker, S., Zahoor, A., Schramm, T., Link, H., and Klamt, S. (2019). "Broadening the scope of enforced ATP wasting as a tool for metabolic engineering in Escherichia coli". In: *Biotechnology Journal* 14.9.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). "Trimmomatic: a flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15, pp. 2114–2120.

Bolouri, H. and Davidson, E. H. (2002). "Modeling transcriptional regulatory networks". In: *BioEssays* 24.12, pp. 1118–1129.

Brandon, M., Howard, B., Lawrence, C., and Laubenbacher, R. (2015). "Iron acquisition and oxidative stress response in Aspergillus fumigatus". In: *BMC Systems Biology* 9.1, p. 19.

Bremer, H. and Dennis, P. (2008). "Feedback control of ribosome function in Escherichia coli". In: *Biochimie* 90.3, pp. 493–499.

Bremer, H. and Dennis, P. P. (1996a). "Escherichia coli and Salmonella: cellular and molecular biology". In: *Washington (DC): American Society for Microbiology. Chapter, Modulation of chemical composition and other parameters of the cell by growth rate*, pp. 1553–1569.

Bremer, H., Dennis, P. P., et al. (1996b). "Modulation of chemical composition and other parameters of the cell by growth rate". In: *Escherichia coli and Salmonella: cellular and molecular biology* 2.2, pp. 1553–69.

Breuer, M., Lakehal, D., and Rodi, W. (1996). "Flow around a surface mounted cubical obstacle: comparison of les and rans-results". In: *Computation of Three-Dimensional Complex Flows*. Springer, pp. 22–30.

Brilman, D. W. F. (2000). "Mass transfer and chemical reaction in gas-liquid-liquid systems."

Brilman, D. W. F., Swaaij, W. P. M. van, and Versteeg, G. (1998). "A one-dimensional instationary heterogeneous mass transfer model for gas absorption in multiphase systems". In: *Chemical engineering and processing: process intensification* 37.6, pp. 471–488.

Brognaux, A., Han, S., Sørensen, S. J., Lebeau, F., Thonart, P., and Delvigne, F. (2013). "A low-cost, multiplexable, automated flow cytometry procedure for the characterization of microbial stress dynamics in bioreactors". In: *Microbial Cell Factories* 12.1, p. 100.

Brown, D. R., Barton, G., Pan, Z., Buck, M., and Wigneshweraraj, S. (2014). "Nitrogen stress response and stringent response are coupled in Escherichia coli". In: *Nature communications* 5.1, pp. 1–8.

Buchholz, J., Graf, M., Freund, A., Busche, T., Kalinowski, J., Blombach, B., and Takors, R. (2014). "CO2/HCO3- perturbations of simulated large scale gradients in a scale-down device cause fast transcriptional responses in Corynebacterium glutamicum". In: *Applied Microbiology and Biotechnology* 98.20, pp. 8563–8572.

Busby, S. (1994). "Promoter structure, promoter recognition, and transcription activation in prokaryotes". In: *Cell* 79.5, pp. 743–746.

Butala, M., Klose, D., Hodnik, V., Rems, A., Podlesek, Z., Klare, J. P., Anderluh, G., Busby, S. J., Steinhoff, H.-J., and Žgur-Bertok, D. (2011). "Interconversion between bound and free conformations of LexA orchestrates the bacterial SOS response". In: *Nucleic acids research* 39.15, pp. 6546–6557.

Bylund, F., Castan, A., Mikkola, R., Veide, A., and Larsson, G. (2000). "Influence of scale-up on the quality of recombinant human growth hormone". In: *Biotechnology and Bioengineering* 69.2, pp. 119–128.

Bylund, F., Collet, E., Enfors, S.-O., and Larsson, G. (1998). "Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation". In: *Bioprocess Engineering* 18.3, pp. 171–180.

Bylund, F., Guillard, F., Enfors, S.-O., Trägårdh, C., and Larsson, G. (1999). "Scale down of recombinant protein production: a comparative study of scaling performance". In: *Bioprocess and Biosystems Engineering* 20.5, pp. 377–389.

Carpousis, A. J., Luisi, B. F., and McDowall, K. J. (2009). "Chapter 3 Endonucleolytic Initiation of mRNA Decay in Escherichia coli". In: *Molecular biology of RNA processing and decay in prokaryotes*. Ed. by C. Condon. Vol. 85. Progress in nucleic acid research and molecular biology. Amsterdam: Elsevier, pp. 91–135.

Caspeta, L., Flores, N., Perez, N. O., Bolivar, F., and Ramirez, O. T. (2009). "The effect of heating rate on *Escherichia coli* metabolism, physiological stress, transcriptional response, and production of temperature-induced recombinant protein: A scale-down study". In: *Biotechnology and Bioengineering* 102.2, pp. 468–482.

Cellerino, A. and Sanguanini, M. (2018). *Transcriptome Analysis: Introduction and Examples from the Neurosciences*. Scuola Normale Superiore.

Chandrasekaran, S. and Price, N. D. (2010). "Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*". In: *Proceedings of the National Academy of Sciences* 107.41, pp. 17845–17850.

Chandrasekhar, T., Thangavel, K., and Elayaraja, E. (2011). "Effective Clustering Algorithms for Gene Expression Data". In: 32(4), pp. 25–9.

Chang, D.-E., Smalley, D. J., and Conway, T. (2002). "Gene expression profiling of Escherichia coli growth transitions: an expanded stringent response model". In: *Molecular microbiology* 45.2, pp. 289–306.

Chapman, A. G. and Atkinson, D. E. (1977). "Adenine nucleotide concentrations and turnover rates. Their correlation with biological activity in bacteria and yeast". In: *Advances in microbial physiology* 15, pp. 253–306.

Chapman, A. G., Fall, L., and Atkinson, D. E. (1971). "Adenylate energy charge in Escherichia coli during growth and starvation". In: *Journal of bacteriology* 108.3, pp. 1072–1086.

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., and Reuss, M. (2002). "Dynamic modeling of the central carbon metabolism of *Escherichia coli*". In: *Biotechnology and Bioengineering* 79.1, pp. 53–73.

Chen, H. and Boutros, P. C. (2011). "VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R". In: *BMC bioinformatics* 12.1, pp. 1–7.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X. S. (2015a). "Genome-wide study of mRNA degradation and transcript elongation in E scherichia coli". In: *Molecular systems biology* 11.1, p. 781.

Chen, H., Shiroguchi, K., Ge, H., and Xie, X. S. (2015b). "Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli". In: *Molecular systems biology* 11.1, p. 781.

Chen, K. C., Calzone, L., Csikasz-Nagy, A., Cross, F. R., Novak, B., and Tyson, J. J. (2004). "Integrative analysis of cell cycle control in budding yeast". In: *Molecular Biology of the Cell* 15.8, pp. 3841–3862.

Chou, I.-C. and Voit, E. O. (2009). "Recent developments in parameter estimation and structure identification of biochemical and genomic systems". In: *Mathematical Biosciences* 219.2, pp. 57–83.

Chowdhury, N., Kwan, B. W., and Wood, T. K. (2016). "Persistence increases in the absence of the alarmone guanosine tetraphosphate by reducing cell growth". In: *Scientific reports* 6, p. 20519.

Chubukov, V., Desmarais, J. J., Wang, G., Chan, L. J. G., Baidoo, E. E., Petzold, C. J., Keasling, J. D., and Mukhopadhyay, A. (2017). "Engineering glucose metabolism of Escherichia coli under nitrogen starvation". In: *NPJ systems biology and applications* 3.1, pp. 1–7.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., et al. (2016). "A survey of best practices for RNA-seq data analysis". In: *Genome Biology* 17.1, p. 13.

Coroneo, M., Montante, G., Paglianti, A., and Magelli, F. (2011). "CFD prediction of fluid flow and mixing in stirred tanks: Numerical issues about the RANS simulations". In: *Computers & Chemical Engineering* 35.10, pp. 1959–1968.

Corson, D., Jaiman, R., and Shakib, F. (2009). "Industrial application of RANS modelling: capabilities and needs". In: *International journal of Computational Fluid dynamics* 23.4, pp. 337–347.

Cortés, J. T., Flores, N., Bolivar, F., Lara, A. R., and Ramirez, O. T. (2016). "Physiological effects of pH gradients on Escherichia coli during plasmid DNA production". In: *Biotechnology and bioengineering* 113.3, pp. 598–611.

Costessi, A., Bogert, B. van den, May, A., Ver Loren van Themaat, E., Roubos, J. A., Kolkman, M. A., Butler, D., and Pirovano, W. (2018). "Novel sequencing technologies to support industrial biotechnology". In: *FEMS microbiology letters* 365.16.

Cox, R. A. (2004). "Quantitative relationships for specific growth rates and macromolecular compositions of Mycobacterium tuberculosis, Streptomyces coelicolor A3 (2) and Escherichia coli B/r: an integrative theoretical approach". In: *Microbiology* 150.5, pp. 1413–1426.

Cuvelier, M.-E., Soto, P., Courtois, F., Broyart, B., and Bonazzi, C. (2017). "Oxygen solubility measured in aqueous or oily media by a method using a non-invasive sensor". In: *Food control* 73, pp. 1466–1473.

D'haeseleer, P. (2005). "How does gene expression clustering work?" In: *Nature biotechnology* 23.12, pp. 1499–1501.

Dalebroux, Z. D. and Swanson, M. S. (2012). "ppGpp: magic beyond RNA polymerase". In: *Nature Reviews Microbiology* 10.3, pp. 203–212.

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). "Mauve: multiple alignment of conserved genomic sequence with rearrangements". In: *Genome research* 14.7, pp. 1394–1403.

Datta, S., Costantino, N., and Court, D. L. (2006). "A set of recombineering plasmids for gram-negative bacteria". In: *Gene* 379, pp. 109–115.

Davidich, M. and Bornholdt, S. (2008). "The transition from differential equations to Boolean networks: a case study in simplifying a regulatory network model". In: *Journal of Theoretical Biology* 255.3, pp. 269–277.

De Anda, R., Lara, A. R., Hernandez, V., Hernandez-Montalvo, V., Gosset, G., Bolivar, F., and Ramirez, O. T. (2006). "Replacement of the glucose phosphotransferase transport system by galactose permease reduces acetate accumulation and improves process performance of Escherichia coli for recombinant protein production without impairment of growth rate". In: *Metabolic engineering* 8.3, pp. 281–290.

Delafosse, A., Collignon, M.-L., Calvo, S., Delvigne, F., Crine, M., Thonart, P., and Toye, D. (2014). "CFD-based compartment model for description of mixing in bioreactors". In: *Chemical Engineering Science* 106, pp. 76–85.

Delvigne, F., Boxus, M., Ingels, S., and Thonart, P. (2009). "Bioreactor mixing efficiency modulates the activity of a prpoS:: GFP reporter gene in E. coli". In: *Microbial cell factories* 8.1, pp. 1–17.

Delvigne, F., Destain, J., and Thonart, P. (2006). "A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models". In: *Biochemical Engineering Journal* 28.3, pp. 256–268.

Delvigne, F., Takors, R., Mudde, R., Gulik, W., and Noorman, H. (2017). "Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond". In: *Microbial Biotechnology*, 10(5), pp. 1267–1274.

Dewan, A., Buwa, V., and Durst, F. (2006). "Performance optimizations of grid disc impellers for mixing of single-phase flows in a stirred vessel". In: *Chemical Engineering Research and Design* 84.8, pp. 691–702.

Dickey, D. S. and DS, D. (1976). "Dimensional analysis for fluid agitation systems". In.

Doucette, C. D., Schwab, D. J., Wingreen, N. S., and Rabinowitz, J. D. (2011). "α-Ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition". In: *Nature chemical biology* 7.12, p. 894.

Drăghici, S. (2012). *Statistics and data analysis for microarrays using R and Bioconductor*. Boca Raton, FL: CRC Press.

Dumont, E. and Delmas, H. (2003). "Mass transfer enhancement of gas absorption in oil-in-water systems: a review". In: *Chemical Engineering and Processing: Process Intensification* 42.6, pp. 419–438.

Durfee, T., Hansen, A.-M., Zhi, H., Blattner, F. R., and Jin, D. J. (2008). "Transcription profiling of the stringent response in Escherichia coli". In: *Journal of bacteriology* 190.3, pp. 1084–1096.

Edgar, R., Domrachev, M., and Lash, A. E. (2002). "Gene Expression Omnibus: NCBI gene expression and hybridization array data repository". In: *Nucleic acids research* 30.1, pp. 207–210.

Egli, T. (1991). “On multiple-nutrient-limited growth of microorganisms, with special reference to dual limitation by carbon and nitrogen substrates”. In: *Antonie van Leeuwenhoek* 60.3-4, pp. 225–234.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). “Cluster analysis and display of genome-wide expression patterns”. In: *Proceedings of the National Academy of Sciences* 95.25, pp. 14863–14868.

Eiteman, M. A. and Altman, E. (2006). “Overcoming acetate in Escherichia coli recombinant protein fermentations”. In: *Trends in biotechnology* 24.11, pp. 530–536.

Elkins, P. A., Watts, J. M., Zalacain, M., Thiel, A. van, Vitazka, P. R., Redlak, M., Andraos-Selim, C., Rastinejad, F., and Holmes, W. M. (2003). “Insights into catalysis by a knotted TrmD tRNA methyltransferase”. In: *Journal of molecular biology* 333.5, pp. 931–949.

Enfors, S.-O., Jahic, M., Rozkov, A., Xu, B., Hecker, M., Jürgen, B., Krüger, E., Schweder, T., Hamer, G., O’beirne, D., et al. (2001). “Physiological responses to mixing in large scale bioreactors”. In: *Journal of Biotechnology* 85.2, pp. 175–185.

Erickson, D. W., Schink, S. J., Patsalo, V., Williamson, J. R., Gerland, U., and Hwa, T. (2017). “A global resource allocation strategy governs growth transition kinetics of *Escherichia coli*”. In: *Nature* 551.7678, pp. 119–123.

Escherich, T. (1988). “The intestinal bacteria of the neonate and breast-fed infant”. In: *Clinical Infectious Diseases* 10.6, pp. 1220–1225.

Ewing, B. and Green, P. (1998). “Base-calling of automated sequencer traces using phred. II. Error probabilities”. In: *Genome research* 8.3, pp. 186–194.

Ewing, B., Hillier, L., Wendl, M. C., and Green, P. (1998). “Base-calling of automated sequencer traces usingPhred. I. Accuracy assessment”. In: *Genome research* 8.3, pp. 175–185.

Faiß, J. (2020). “Untersuchung des Einflusses unterschiedlicher Rührer und Medien auf den Sauerstoff Massentransfer”.

Ferrer-Miralles, N., Domingo-Espin, J., Corchero, J. L., Vazquez, E., and Villaverde, A. (2009). “Microbial factories for recombinant pharmaceuticals”. In: *Microbial cell factories* 8.1, pp. 1–8.

Fichtel, K. (2019). “Effekt unterschiedlicher Energieeinträge und Rührorgane auf die Tropfengrößenverteilung von Öl in Wasser”.

Figueira, R., Brown, D. R., Ferreira, D., Eldridge, M. J., Burchell, L., Pan, Z., Helaine, S., and Wigneshweraraj, S. (2015). “Adaptation to sustained nitrogen starvation by Escherichia coli requires the eukaryote-like serine/threonine kinase YeaG”. In: *Scientific reports* 5, p. 17524.

Fluent, A. (2009). "12.0 Theory Guide". In: *Ansys Inc* 5.

Fonville, N. C., Bates, D., Hastings, P., Hanawalt, P. C., and Rosenberg, S. M. (2010). "Role of RecA and the SOS response in thymineless death in Escherichia coli". In: *PLoS Genet* 6.3.

Freddolino, P. L., Amini, S., and Tavazoie, S. (2012). "Newly identified genetic variations in common Escherichia coli MG1655 stock cultures". In: *Journal of bacteriology* 194.2, pp. 303–306.

Fröhling, M. and Hiete, M. (2020). "Sustainability and life cycle assessment in industrial biotechnology: A review of current approaches and future needs". In: *Sustainability and Life Cycle Assessment in Industrial Biotechnology*, pp. 143–203.

Gaca, A. O., Kudrin, P., Colomer-Winter, C., Beljantseva, J., Liu, K., Anderson, B., Wang, J. D., Rejman, D., Potrykus, K., Cashel, M., et al. (2015). "From (p) ppGpp to (pp) pGpp: characterization of regulatory effects of pGpp synthesized by the small alarmone synthetase of Enterococcus faecalis". In: *Journal of bacteriology* 197.18, pp. 2908–2919.

Gallant, J., Erlich, H., Hall, B., and Laffler, T. (1970). "Analysis of the RC function". In: *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 35. Cold Spring Harbor Laboratory Press, pp. 397–405.

Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muniz-Rascado, L., Garcia-Sotelo, J. S., Alquicira-Hernandez, K., Martinez-Flores, I., Pannier, L., Castro-Mondragon, J. A., et al. (2016). "RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond". In: *Nucleic Acids Research* 44.D1, pp. D133–D143.

Garcia-Ochoa, F. and Gomez, E. (2009). "Bioreactor scale-up and oxygen transfer rate in microbial processes: an overview". In: *Biotechnology advances* 27.2, pp. 153–176.

Gennemark, P. and Wedelin, D. (2009). "Benchmarks for identification of ordinary differential equations from time series data". In: *Bioinformatics* 25.6, pp. 780–786.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). "Bioconductor: open software development for computational biology and bioinformatics". In: *Genome biology* 5.10, p. 80.

George, S., Larsson, G., and Enfors, S.-O. (1993). "A scale-down two-compartment reactor with controlled substrate oscillations: metabolic response of *Saccharomyces cerevisiae*". In: *Bioprocess and Biosystems Engineering* 9.6, pp. 249–257.

Gerdes, K. and Maisonneuve, E. (2012). "Bacterial persistence and toxin-antitoxin loci". In: *Annual review of microbiology* 66, pp. 103–123.

Gerding, M. A., Ogata, Y., Pecora, N. D., Niki, H., and De Boer, P. A. (2007). "The trans-envelope Tol–Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E. coli". In: *Molecular microbiology* 63.4, pp. 1008–1025.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). "Enzymatic assembly of DNA molecules up to several hundred kilobases". In: *Nature methods* 6.5, pp. 343–345.

Gillespie, D. T. (1976). "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* 22.4, pp. 403–434.

Gillespie, D. T. (1977). "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25, pp. 2340–2361.

Glick, B. R. (1995). "Metabolic load and heterologous gene expression". In: *Biotechnology advances* 13.2, pp. 247–261.

Goldberg, A. L. and St John, A. C. (1976). "Intracellular protein degradation in mammalian and bacterial cells: Part 2". In: *Annual review of biochemistry* 45, pp. 747–803.

Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998). "The Functional and Regulatory Roles of Sigma Factors in Transcription". In: *Cold Spring Harbor Symposia on Quantitative Biology* 63.0, pp. 141–156.

Gualerzi, C. O. and Pon, C. L. (2015). "Initiation of mRNA translation in bacteria: structural and dynamic aspects". In: *Cellular and molecular life sciences : CMLS* 72.22, pp. 4341–4367.

Guido, N. J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C. R., Elston, T. C., and Collins, J. (2006). "A bottom-up approach to gene regulation". In: *Nature* 439.7078, p. 856.

Gunyol, O. and Mudde, R. (2009). "Computational study of hydrodynamics of a standard stirred tank reactor and a large-scale multi-impeller fermenter". In: *International Journal for Multiscale Computational Engineering* 7.6.

Gupta, S., Singh, S. N., and Kumar, D. (2016). "Clustering Methods Applied for Gene Expression Data: A Study". In: *2016 Second International Conference on Computational Intelligence & Communication Technology: CICT 2016 : Ghaziabad, India, 12-13 February 2016 : proceedings*. Ed. by Technology, International Conference on Computational Intelligence &. Communication. Piscataway, NJ: IEEE, pp. 724–728.

Gur, E., Biran, D., and Ron, E. Z. (2011). "Regulated proteolysis in Gram-negative bacteria–how and when?" In: *Nature reviews. Microbiology* 9.12, pp. 839–848.

Halkidi, M., Batistakis, Y., and Vazirgiannis, M. (2001). In: *Journal of Intelligent Information Systems* 17.2/3, pp. 107–145.

Hanahan, D. (1983). "Studies on transformation of Escherichia coli with plasmids". In: *Journal of molecular biology* 166.4, pp. 557–580.

Handl, J., Knowles, J., and Kell, D. B. (2005). "Computational cluster validation in post-genomic data analysis". In: *Bioinformatics (Oxford, England)* 21.15, pp. 3201–3212.

Hara, K. Y. and Kondo, A. (2015). "ATP regulation in bioproduction". In: *Microbial cell factories* 14.1, p. 198.

Hardiman, T., Lemuth, K., Keller, M. A., Reuss, M., and Siemann-Herzberg, M. (2007). "Topology of the global regulatory network of carbon limitation in *Escherichia coli*". In: *Journal of Biotechnology* 132.4, pp. 359–374.

Hardiman, T., Lemuth, K., Siemann-Herzberg, M., and Reuss, M. (2009). "Dynamic Modeling of the Central Metabolism of *E. coli*–Linking Metabolite and Regulatory Networks". In: *Systems Biology and Biotechnology of Escherichia coli*. Springer, pp. 209–235.

Haringa, C., Deshmukh, A. T., Mudde, R. F., and Noorman, H. J. (2017). "Euler-Lagrange analysis towards representative down-scaling of a 22 cubic meters aerobic *S. cerevisiae* fermentation". In: *Chemical Engineering Science*, 16(7), pp. 652–663.

Haringa, C., Tang, W., Deshmukh, A. T., Xia, J., Reuss, M., Heijnen, J. J., Mudde, R. F., and Noorman, H. J. (2016). "Euler-Lagrange computational fluid dynamics for (bio) reactor scale down: An analysis of organism lifelines". In: *Engineering in Life Sciences* 16.7, pp. 652–663.

Haringa, C., Tang, W., Wang, G., Deshmukh, A. T., Winden, W. A. van, Chu, J., Gulik, W. M. van, Heijnen, J. J., Mudde, R. F., and Noorman, H. J. (2018). "Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: towards rational scale-down and design optimization". In: *Chemical Engineering Science* 175, pp. 12–24.

Haringa, C., Vandewijer, R., and Mudde, R. F. (2018a). "Inter-compartment interaction in multi-impeller mixing: Part I. Experiments and multiple reference frame CFD". In: *Chemical Engineering Research and Design* 136, pp. 870–885.

Haringa, C., Vandewijer, R., and Mudde, R. F. (2018b). "Inter-compartment interaction in multi-impeller mixing. Part II. Experiments, sliding mesh and large Eddy simulations". In: *Chemical Engineering Research and Design* 136, pp. 886–899.

Hauryliuk, V., Atkinson, G. C., Murakami, K. S., Tenson, T., and Gerdes, K. (2015). "Recent functional insights into the role of (p) ppGpp in bacterial physiology". In: *Nature Reviews Microbiology* 13.5, pp. 298–309.

Hayashi, K., Morooka, N., Yamamoto, Y., Fujita, K., Isono, K., Choi, S., Ohtsubo, E., Baba, T., Wanner, B. L., Mori, H., et al. (2006). "Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110". In: *Molecular systems biology* 2.1, pp. 2006–0007.

Heimann, N. (2017). "Effect of surface tension, gas composition and viscosity on bubble size distribution in stirred tank reactors".

Heins, A.-L., Lencastre Fernandes, R., Gernaey, K. V., and Lantz, A. E. (2015). "Experimental and in silico investigation of population heterogeneity in continuous *Saccharomyces cerevisiae* scale-down fermentation in a two-compartment setup". In: *Journal of Chemical Technology and Biotechnology* 90.2, pp. 324–340.

Heinzle, E., Biwer, A. P., and Cooney, C. L. (2007). *Development of sustainable bioprocesses: modeling and assessment*. John Wiley & Sons.

Helmann, J. D. and Chamberlin, M. J. (1988). "Structure and function of bacterial sigma factors". In: *Annual review of biochemistry* 57, pp. 839–872.

Henry, C. L., Dalton, C. N., Scruton, L., and Craig, V. S. (2007). "Ion-specific coalescence of bubbles in mixed electrolyte solutions". In: *The Journal of Physical Chemistry C* 111.2, pp. 1015–1023.

Heyer, L. J. (1999). "Exploring Expression Data: Identification and Analysis of Coexpressed Genes". In: *Genome Research* 9.11, pp. 1106–1115.

Heyland, J., Blank, L. M., and Schmid, A. (2011). "Quantification of metabolic limitations during recombinant protein production in Escherichia coli". In: *Journal of Biotechnology* 155.2, pp. 178–184.

Higbie, R. (1935). "The rate of absorption of a pure gas into a still liquid during short periods of exposure". In: *Trans. AIChE* 31, pp. 365–389.

Holland, F. and Bragg, R. (1995). *Fluid flow for chemical and process engineers*. Elsevier.

Hopkins, D. J., Betenbaugh, M. J., and Dhurjati, P. (1987). "Effects of dissolved oxygen shock on the stability of recombinant Escherichia coli containing plasmid pKN401". In: *Biotechnology and bioengineering* 29.1, pp. 85–91.

Hua, Q., Yang, C., Oshima, T., Mori, H., and Shimizu, K. (2004). "Analysis of gene expression in Escherichia coli in response to changes of growth-limiting nutrient in chemostat cultures". In: *Applied and environmental microbiology* 70.4, pp. 2354–2366.

Hui, M. P., Foley, P. L., and Belasco, J. G. (2014). "Messenger RNA Degradation in Bacterial Cells". In: *Annual review of genetics* 48, pp. 537–559.

Huisman, O., D'Ari, R., and Gottesman, S. (1984). "Cell-division control in Escherichia coli: specific induction of the SOS function SfiA protein is sufficient to block septation". In: *Proceedings of the National Academy of Sciences* 81.14, pp. 4490–4494.

Hunkapiller, T., Kaiser, R., Koop, B., and Hood, L. (1991). "Large-scale and automated DNA sequence determination". In: *Science* 254.5028, pp. 59–67.

Iglesias, F. and Kastner, W. (2013). "Analysis of similarity measures in times series clustering for the discovery of building energy patterns". In: *Energies* 6.2, pp. 579–597.

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., et al. (2007). "Multiple high-throughput analyses monitor the response of *E. coli* to perturbations". In: *Science* 316.5824, pp. 593–597.

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). "Data clustering: a review". In: *ACM Computing Surveys* 31.3, pp. 264–323.

Jain, A. K. (2010). "Data clustering: 50 years beyond K-means". In: *Pattern recognition letters* 31.8, pp. 651–666.

Jain, A. K. and Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice-Hall.

Jarmander, J., Belotserkovsky, J., Sjöberg, G., Guevara-Martinez, M., Pérez-Zabaleta, M., Quillaguamán, J., and Larsson, G. (2015). "Cultivation strategies for production of (R)-3-hydroxybutyric acid from simultaneous consumption of glucose, xylose and arabinose by Escherichia coli". In: *Microbial cell factories* 14.1, p. 51.

Jenal, U. and Hengge-Aronis, R. (2003). "Regulation by proteolysis in bacterial cells". In: *Current Opinion in Microbiology* 6.2, pp. 163–172.

Jiang, D., Tang, C., and Zhang, A. (2004). "Cluster analysis for gene expression data: a survey". In: *IEEE Transactions on Knowledge and Data Engineering* 16.11, pp. 1370–1386.

Jishage, M., Iwata, A., Ueda, S., and Ishihama, A. (1996). "Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions." In: *Journal of bacteriology* 178.18, pp. 5447–5451.

Johansson, J., Balsalobre, C., Wang, S.-Y., Urbonaviciene, J., Jin, D. J., Sondén, B., and Uhlin, B. E. (2000). "Nucleoid proteins stimulate stringently controlled bacterial promoters: a link between the cAMP-CRP and the (p) ppGpp regulons in Escherichia coli". In: *Cell* 102.4, pp. 475–485.

Jolliffe, I. T. and Cadima, J. (2016). "Principal component analysis: a review and recent developments". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2065.

Jonge, L. P. de, Buijs, N. A., Pierick, A. ten, Deshmukh, A., Zhao, Z., Kiel, J. A., Heijnen, J. J., and Gulik, W. M. van (2011). "Scale-down of penicillin production in Penicillium chrysogenum". In: *Biotechnology journal* 6.8, pp. 944–958.

Joseleau-Petit, D., Vinella, D., and D'Ari, R. (1999). "Metabolic alarms and cell division inescherichia coli". In: *Journal of bacteriology* 181.1, pp. 9–14.

Junker, B. H. (2004). "Scale-up methodologies for Escherichia coli and yeast fermentation processes". In: *Journal of bioscience and bioengineering* 97.6, pp. 347–364.

Junker, B., Stanik, M., Barna, C., Salmon, P., Paul, E., and Buckland, B. (1998). "Influence of impeller type on power input in fermentation vessels". In: *Bioprocess Engineering* 18.6, pp. 401–412.

Junne, S., Klingner, A., Kabisch, J., Schweder, T., and Neubauer, P. (2011). "A two-compartment bioreactor system made of commercial parts for bioprocess scale-down studies: Impact of oscillations on Bacillus subtilis fed-batch cultivations". In: *Biotechnology Journal* 6.8, pp. 1009–1017.

Kaleta, C., Schäuble, S., Rinas, U., and Schuster, S. (2013). "Metabolic costs of amino acid and protein production in *Escherichia coli*". In: *Biotechnology Journal* 8.9, pp. 1105–1114.

Kanjee, U., Gutsche, I., Alexopoulos, E., Zhao, B., El Bakkouri, M., Thibault, G., Liu, K., Ramachandran, S., Snider, J., Pai, E. F., et al. (2011). "Linkage between the bacterial acid stress and stringent responses: the structure of the inducible lysine decarboxylase". In: *The EMBO journal* 30.5, pp. 931–944.

Kanjee, U., Ogata, K., and Houry, W. A. (2012). "Direct binding targets of the stringent response alarmone (p) ppGpp". In: *Molecular microbiology* 85.6, pp. 1029–1043.

Karcagi, I., Draskovits, G., Umenhoffer, K., Fekete, G., Kovacs, K., Mehi, O., Baliko, G., Szappanos, B., Györfy, Z., Feher, T., et al. (2016). "Indispensability of horizontally transferred genes and its impact on bacterial genome streamlining". In: *Molecular biology and evolution* 33.5, pp. 1257–1269.

Karlebach, G. and Shamir, R. (2008). "Modelling and analysis of gene regulatory networks". In: *Nature Reviews Molecular Cell Biology* 9.10, pp. 770–780.

Karpinets, T. V., Greenwood, D. J., Sams, C. E., and Ammons, J. T. (2006). "RNA: protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis". In: *BMC biology* 4.1, p. 30.

Kauffman, S., Peterson, C., Samuelsson, B., and Troein, C. (2003). "Random Boolean network models and the yeast transcriptional network". In: *Proceedings of the National Academy of Sciences* 100.25, pp. 14796–14799.

Kaur, R., Ramakrishna, M., and Nigam, K. (2007). "Role of dispersed phase in gas-liquid reactions: A review". In: *Reviews in Chemical Engineering* 23.3-4, pp. 247–300.

Kelly, W. J. (2008). "Using computational fluid dynamics to characterize and improve bioreactor performance". In: *Biotechnology and Applied Biochemistry* 49.4, pp. 225–238.

Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., Fulcher, C., Huerta, A. M., Kothari, A., Krummenacker, M., et al. (2013). "EcoCyc: fusing model organism databases with systems biology". In: *Nucleic Acids Research* 41.D1, pp. D605–D612.

Khlebnikov, A., Risa, O., Skaug, T., Carrier, T. A., and Keasling, J. (2000). "Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture". In: *Journal of Bacteriology* 182.24, pp. 7029–7034.

Kibbe, W. A. (2007). "OligoCalc: an online oligonucleotide properties calculator". In: *Nucleic acids research* 35.suppl_2, W43–W46.

Kitano, H. (2000). "Perspectives on systems biology". In: *New Generation Computing* 18.3, pp. 199–216.

Klipp, E., Herwig, R., Kowald, A., Wierling, C., and Lehrach, H. (2008). *Systems biology in practice: concepts, implementation and application*. John Wiley & Sons.

Klipp, E., Nordlander, B., Krüger, R., Gennemark, P., and Hohmann, S. (2005). "Integrative model of the response of yeast to osmotic shock". In: *Nature Biotechnology* 23.8, p. 975.

Koebmann, B. J., Westerhoff, H. V., Snoep, J. L., Nilsson, D., and Jensen, P. R. (2002). "The glycolytic flux in Escherichia coli is controlled by the demand for ATP". In: *Journal of bacteriology* 184.14, pp. 3909–3916.

Korch, S. B., Malhotra, V., Contreras, H., and Clark-Curtiss, J. E. (2015). "The Mycobacterium tuberculosis relBE toxin: antitoxin genes are stress-responsive modules that regulate growth through translation inhibition". In: *Journal of Microbiology* 53.11, pp. 783–795.

Kremling, A., Heermann, R., Centler, F., Jung, K., and Gilles, E. (2004). "Analysis of two-component signal transduction by mathematical modeling using the KdpD/KdpE system of *Escherichia coli*". In: *Biosystems* 78.1, pp. 23–37.

Kremling, A., Bettenbrock, K., and Gilles, E. D. (2007). "Analysis of global control of *Escherichia coli* carbohydrate uptake". In: *BMC Systems Biology* 1.1, p. 42.

Krist, S. (2013). "Distelöl". In: *Lexikon der pflanzlichen Fette und Öle*. Springer, pp. 205–213.

KRÜSS (2020). "Pressure characteristic for the bubble pressure measurement, position of pressure maximum". In: URL: https://www.kruss-scientific.com/en/know-how/glossary/bubble-pressure-tensiometer,%20downloaded:%2028.01.21.

Kurokawa, M., Seno, S., Matsuda, H., and Ying, B.-W. (2016). "Correlation between genome reduction and bacterial growth". In: *Dna Research* 23.6, pp. 517–525.

Kuschel, M., Siebler, F., and Takors, R. (2017). "Lagrangian Trajectories to Predict the Formation of Population Heterogeneity in Large-Scale Bioreactors". In: *Bioengineering* 4.2, p. 27.

Kuschel, M. and Takors, R. (2020). "Simulated oxygen and glucose gradients as a prerequisite for predicting industrial scale performance a priori". In: *Biotechnology and Bioengineering* 117.9, pp. 2760–2770.

Laalami, S., Zig, L., and Putzer, H. (2014). "Initiation of mRNA decay in bacteria". In: *Cellular and molecular life sciences : CMLS* 71.10, pp. 1799–1828.

Lam, H. Y., Clark, M. J., Chen, R., Chen, R., Natsoulis, G., O'huallachain, M., Dewey, F. E., Habegger, L., Ashley, E. A., Gerstein, M. B., et al. (2012). "Performance comparison of whole-genome sequencing platforms". In: *Nature biotechnology* 30.1, pp. 78–82.

Langmead, B. and Salzberg, S. L. (2012). "Fast gapped-read alignment with Bowtie 2". In: *Nature methods* 9.4, p. 357.

Lapin, A., Klann, M., and Reuss, M. (2010). "Multi-scale spatio-temporal modeling: lifelines of microorganisms in bioreactors and tracking molecules in cells". In: *Biosystems Engineering II*. Springer, pp. 23–43.

Lapin, A., Müller, D., and Reuss, M. (2004). "Dynamic Behavior of Microbial Populations in Stirred Bioreactors Simulated with Euler- Lagrange Methods: Traveling along the Lifelines of Single Cells". In: *Industrial & Engineering Chemistry Research* 43.16, pp. 4647–4656.

Lapin, A., Schmid, J., and Reuss, M. (2006). "Modeling the dynamics of *E. coli* populations in the three-dimensional turbulent field of a stirred-tank bioreactor—A structured–segregated approach". In: *Chemical Engineering Science* 61.14, pp. 4783–4797.

Lara, A., Galindo, E., Ramirez, O., and Palomares, L. (2006). "Living with heterogeneities in bioreactors". In: *Molecular Biotechnology* 34.3, pp. 355–381.

Lara, A., Leal, L., Flores, N., Gosset, G., Bolivar, F., and Ramirez, O. (2006). "Transcriptional and metabolic response of recombinant *Escherichia coli* to spatial dissolved oxygen tension gradients simulated in a scale-down system". In: *Biotechnology and Bioengineering* 93.2, pp. 372–385.

Lara, A., Palomares, L., and Ramirez, O. (2016). "Scale-down: Simulating large-scale cultures in the laboratory". In: *Industrial Biotechnology: Products and Processes*, pp. 55–79.

Lara, A. R., Taymaz-Nikerel, H., Mashego, M. R., Gulik, W. M. van, Heijnen, J. J., Ramirez, O. T., and Winden, W. A. van (2009). "Fast dynamic response of the fermentative

metabolism of *Escherichia coli* to aerobic and anaerobic glucose pulses". In: *Biotechnology and Bioengineering* 104.6, pp. 1153–1161.

Larsson, G. and Enfors, S.-O. (1988). "Studies of insufficient mixing in bioreactors: Effects of limiting oxygen concentrations and short term oxygen starvation on Penicillium chrysogenum". In: *Bioprocess and Biosystems Engineering* 3.3, pp. 123–127.

Larsson, G., Törnkvist, M., Wernersson, E. S., Trägårdh, C., Noorman, H., and Enfors, S.-O. (1996). "Substrate gradients in bioreactors: origin and consequences". In: *Bioprocess and Biosystems Engineering* 14.6, pp. 281–289.

Laszlo, M. and Mukherjee, S. (2007). "A genetic algorithm that exchanges neighboring centers for k-means clustering". In: *Pattern Recognition Letters* 28.16, pp. 2359–2366.

Launder, B. E. and Spalding, D. B. (1972). *Lectures in mathematical models of turbulence*. Academic Press.

Lee, C. M., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Gonzalez, J. N., Hinrichs, A. S., Lee, B. T., Nassar, L. R., Powell, C. C., et al. (2020). "UCSC Genome Browser enters 20th year". In: *Nucleic acids research* 48.D1, pp. D756–D761.

Lee, K. H., Park, J. H., Kim, T. Y., Kim, H. U., and Lee, S. Y. (2007). "Systems metabolic engineering of *Escherichia coli* for L-threonine production". In: *Molecular Systems Biology* 3.1, p. 149.

Lee, S., Kim, M. H., Kang, B. S., Kim, J.-S., Kim, G.-H., Kim, Y.-G., and Kim, K. J. (2008). "Crystal structure of Escherichia coli MazG, the regulator of nutritional stress response". In: *Journal of Biological Chemistry* 283.22, pp. 15232–15240.

Leimbach, A., Hacker, J., and Dobrindt, U. (2013). "E. coli as an all-rounder: the thin line between commensalism and pathogenicity". In: *Between pathogenicity and commensalism*. Springer, pp. 3–32.

Leisch, F. (2006). "A toolbox for k-centroids cluster analysis". In: *Computational statistics & data analysis* 51.2, pp. 526–544.

Lemuth, K., Hardiman, T., Winter, S., Pfeiffer, D., Keller, M., Lange, S., Reuss, M., Schmid, R., and Siemann-Herzberg, M. (2008). "Global transcription and metabolic flux analysis of *Escherichia coli* in glucose-limited fed-batch cultivations". In: *Applied and Environmental Microbiology* 74.22, pp. 7002–7015.

Li, B., Ruotti, V., Stewart, R. M., Thomson, J. A., and Dewey, C. N. (2010). "RNA-Seq gene expression estimation with read mapping uncertainty". In: *Bioinformatics* 26.4, pp. 493–500.

Li, N., Han, Z., O'Donnell, T. J., Kurasaki, R., Kajihara, L., Williams, P. G., Tang, Y., and Su, W. W. (2020). "Production and excretion of astaxanthin by engineered Y arrowia

lipolytica using plant oil as both the carbon source and the biocompatible extractant". In: *Applied Microbiology and Biotechnology* 104.16, pp. 6977–6989.

Li, X.-t., Thomason, L. C., Sawitzke, J. A., Costantino, N., and Court, D. L. (2013). "Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in Escherichia coli". In: *Nucleic acids research* 41.22, e204–e204.

Li, Y. and Wang, G. (2016). "Strategies of isoprenoids production in engineered bacteria". In: *Journal of applied microbiology* 121.4, pp. 932–940.

Li, Z., Li, P., Krishnan, A., and Liu, J. (2011). "Large-scale dynamic gene regulatory network inference combining differential equation models with local dynamic Bayesian network analysis". In: *Bioinformatics* 27.19, pp. 2686–2691.

Liao, Y., Smyth, G. K., and Shi, W. (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features". In: *Bioinformatics* 30.7, pp. 923–930.

Lieder, S., Nikel, P. I., Lorenzo, V. de, and Takors, R. (2015). "Genome reduction boosts heterologous gene expression in Pseudomonas putida". In: *Microbial cell factories* 14.1, pp. 1–14.

Limberg, M. H., Pooth, V., Wiechert, W., and Oldiges, M. (2016). "Plug flow versus stirred tank reactor flow characteristics in two-compartment scale-down bioreactor: Setup-specific influence on the metabolic phenotype and bioprocess performance of Corynebacterium glutamicum". In: *Engineering in Life Sciences* 16.7, pp. 610–619.

Linnarsson, S. (2010). "Recent advances in DNA sequencing methods–general principles of sample preparation". In: *Experimental cell research* 316.8, pp. 1339–1343.

Liu, X., Yang, S., Wang, F., Dai, X., Yang, Y., and Bai, Z. (2017). "Comparative analysis of the Corynebacterium glutamicum transcriptome in response to changes in dissolved oxygen levels". In: *Journal of Industrial Microbiology & Biotechnology* 44.2, pp. 181–195.

Lloyd, S. (1982). "Least squares quantization in PCM". In: *IEEE Transactions on Information Theory* 28.2, pp. 129–137.

Lo, T.-M., Chng, S. H., Teo, W. S., Cho, H.-S., and Chang, M. W. (2016). "A two-layer gene circuit for decoupling cell growth from metabolite production". In: *Cell systems* 3.2, pp. 133–143.

Lodish, H. (2000). *Molecular cell biology*.

Löffler, M., Simen, J. D., Jäger, G., Schäferhoff, K., Freund, A., and Takors, R. (2016a). "Engineering E. coli for large-scale production - Strategies considering ATP expenses and transcriptional responses". In: *Metabolic engineering* 38, pp. 73–85.

Löffler, M., Simen, J. D., Jäger, G., Schäferhoff, K., Freund, A., and Takors, R. (2016b). "Engineering *E. coli*for large-scale production–Strategies considering ATP expenses and transcriptional responses". In: *Metabolic Engineering* 38, pp. 73–85.

Löffler, M., Simen, J. D., Müller, J., Jäger, G., Laghrami, S., Schäferhoff, K., Freund, A., Takors, R., et al. (2017). "Switching between nitrogen and glucose limitation: Unraveling transcriptional dynamics in *Escherichia coli*". In: *Journal of Biotechnology*, 285: 2–12.

Love, M., Anders, S., and Huber, W. (2014). "Differential analysis of count data–the DESeq2 package". In: *Genome Biol* 15, p. 550.

Lu, Y., Lu, S., Fotouhi, F., Deng, Y., and Brown, S. J. (2004). "Incremental genetic K-means algorithm and its application in gene expression data analysis". In: *BMC bioinformatics* 5, p. 172.

Luo, J. and Gosman, A. (1994). "Prediction of impeller-induced flow in mixing vessels using multiple frames of reference". In.

Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., and Woolf, P. J. (2009). "GAGE: generally applicable gene set enrichment for pathway analysis". In: *BMC bioinformatics* 10.1, p. 161.

Machado, D., Costa, R. S., Rocha, M., Ferreira, E. C., Tidor, B., and Rocha, I. (2011). "Modeling formalisms in Systems Biology". In: *AMB Express* 1, p. 45.

Magnusson, L. U., Farewell, A., and Nyström, T. (2005). "ppGpp: a global regulator in *Escherichia coli*". In: *Trends in Microbiology* 13.5, pp. 236–242.

Mahmoud, S. A. and Chien, P. (2018). "Regulated Proteolysis in Bacteria". In: *Annual review of biochemistry* 87, pp. 677–696.

Mao, X., Ma, Q., Liu, B., Chen, X., Zhang, H., and Xu, Y. (2015). "Revisiting operons: an analysis of the landscape of transcriptional units in E. coli". In: *Bmc Bioinformatics* 16.1, pp. 1–9.

Martinez-Garcia, E., Nikel, P. I., Chavarria, M., and Lorenzo, V. de (2014). "The metabolic cost of flagellar motion in P seudomonas putida KT 2440". In: *Environmental microbiology* 16.1, pp. 291–303.

Maurizi, M. R. (1992a). "Proteases and protein degradation in Escherichia coli". In: *Experientia* 48.2, pp. 178–201.

Maurizi, M. (1992b). "Proteases and protein degradation inEscherichia coli". In: *Experientia* 48.2, pp. 178–201.

McCarthy, J. E. G. and Gualerzi, C. (1990). "Translational control of prokaryotic gene expression". In: *Trends in Genetics* 6, pp. 78–85.

McMillan, J. D. and Wang, D. C. (1987). "Enhanced oxygen transfer using oil-in-water dispersions". In: *Annals of the New York Academy of Sciences* 506, pp. 569–582.

Mcmillan, J. D. and WANG, D. C. (1990). "Mechanisms of oxygen transfer enhancement during submerged cultivation in perfluorochemical-in-water dispersions". In: *Annals of the New York Academy of Sciences* 589, pp. 283–300.

Michalowski, A., Siemann-Herzberg, M., and Takors, R. (2017). "*Escherichia coli* HGT: Engineered for high glucose throughput even under slowly growing or resting conditions". In: *Metabolic Engineering* 40, pp. 93–103.

Mimmack, G. M., Mason, S. J., and Galpin, J. S. (2001). "Choice of Distance Matrices in Cluster Analysis: Defining Regions". In: *Journal of Climate* 14.12, pp. 2790–2797.

Mochizuki, A. (2005). "An analytical study of the number of steady states in gene regulatory networks". In: *Journal of Theoretical Biology* 236.3, pp. 291–310.

Montante, G., Moštěk, M., Jahoda, M., and Magelli, F. (2005). "CFD simulations and experimental validation of homogenisation curves and mixing time in stirred Newtonian and pseudoplastic liquids". In: *Chemical Engineering Science* 60.8-9, pp. 2427–2437.

Morchain, J., Gabelle, J.-C., and Cockx, A. (2014). "A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors". In: *AIChE Journal* 60.1, pp. 27–40.

Moreau, Y., Smet, F. de, Thijs, G., Marchal, K., and Moor, B. de (2002). "Functional bioinformatics of microarray data: from expression to regulation". In: *Proceedings of the IEEE* 90.11, pp. 1722–1743.

Mueller, E., Oh, S., Kavalerchik, E., Kappock, T., Meyer, E., Li, C., Ealick, S., and Stubbe, J. (1999). "Investigation of the ATP binding site of Escherichia coli aminoimidazole ribonucleotide synthetase using affinity labeling and site-directed mutagenesis". In: *Biochemistry* 38.31, pp. 9831–9839.

Murphy, J. (1994). "CFD simulation of flows in stirred tank reactors using a sliding mesh technique". In: *Instn. Chem. Engng. Symp. Ser.* Vol. 136, pp. 341–348.

Murphy, K. C. (2016). "λ Recombination and Recombineering." In: *EcoSal Plus* 7.1.

Murray, D. K. and Bremer, H. (1996). "Control ofspoT-dependent ppGpp Synthesis and Degradation inEscherichia coli". In: *Journal of molecular biology* 259.1, pp. 41–57.

Murray, H. D., Schneider, D. A., and Gourse, R. L. (2003). "Control of rRNA expression by small molecules is dynamic and nonredundant". In: *Molecular Cell* 12.1, pp. 125–134.

Nagpal, A., Jatain, A., and Gaur, D. (2013). "Review based on data clustering algorithms". In: IEEE, pp. 298–303.

Naude, I., Xuereb, C., and Bertrand, J. (1998). "Direct prediction of the flows induced by a propeller in an agitated vessel using an unstructured mesh". In: *The Canadian Journal of Chemical Engineering* 76.3, pp. 631–640.

Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990). *Physiology of the bacterial cell: a molecular approach*. Vol. 20. Sinauer Associates Sunderland, MA.

Neubauer, P., Åhman, M., Törnkvist, M., Larsson, G., and Enfors, S.-O. (1995). "Response of guanosine tetraphosphate to glucose fluctuations in fed-batch cultivations of Escherichia coli". In: *Journal of biotechnology* 43.3, pp. 195–204.

Neubauer, P., Häggström, L., and Enfors, S.-O. (1995). "Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations". In: *Biotechnology and Bioengineering* 47.2, pp. 139–146.

Neubauer, P. and Junne, S. (2010). "Scale-down simulators for metabolic analysis of large-scale bioprocesses". In: *Current Opinion in Biotechnology* 21.1, pp. 114–121.

Neubauer, P. and Junne, S. (2016). "Scale-up and scale-down methodologies for bioreactors". In: *Bioreactors: Design, operation and novel applications*.

Ng, K., Fentiman, N., Lee, K., and Yianneskis, M. (1998). "Assessment of sliding mesh CFD predictions and LDA measurements of the flow in a tank stirred by a Rushton impeller". In: *Chemical Engineering Research and Design* 76.6, pp. 737–747.

Nielsen, J., Nikolajsen, K., and Villadsen, J. (1991). "Structured modeling of a microbial system: I. A theoretical study of lactic acid fermentation". In: *Biotechnology and Bioengineering* 38.1, pp. 1–10.

Nieß, A., Löffler, M., Simen, J. D., and Takors, R. (2017). "Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of *E. coli* cultures in large-scale bioreactors: experimental evidence and mathematical model". In: *Frontiers in Microbiology* 8, p. 1195.

Noguchi, Y., Nakai, Y., Shimba, N., Toyosaki, H., Kawahara, Y., Sugimoto, S., and Suzuki, E.-i. (2004). "The energetic conversion competence of Escherichia coli during aerobic respiration studied by 31P NMR using a circulating fermentation system". In: *Journal of biochemistry* 136.4, pp. 509–515.

Noorman, H. (2011). "An industrial perspective on bioreactor scale-down: what we can learn from combined large-scale bioprocess and model fluid studies". In: *Biotechnology journal* 6.8, pp. 934–943.

Noorman, H. J. and Heijnen, J. J. (2017). "Biochemical engineering's grand adventure". In: *Chemical Engineering Science* 170, pp. 677–693.

Nordheim, A. and Knippers, R., eds. (2018). *Molekulare Genetik: 620 Abbildungen*. 11., unveränderte Auflage. Stuttgart and New York: Thieme.

Notley-McRobb, L., King, T., and Ferenci, T. (2002). "rpoS mutations and loss of general stress resistance in Escherichia coli populations as a consequence of conflict between competing stress responses". In: *Journal of bacteriology* 184.3, pp. 806–811.

Oldshue, J. (1966). "Fermentation mixing scale-up techniques". In: *Biotechnology and Bioengineering* 8.1, pp. 3–24.

Oliveira-Filho, E. R., Silva, J. G., Macedo, M. A. de, Taciro, M. K., Gomez, J. G. C., and Silva, L. F. D. (2019). "Investigating nutrient limitation role on improvement of growth and poly (3-hydroxybutyrate) accumulation by Burkholderia sacchari LMG 19450 from xylose as the sole carbon source". In: *Frontiers in bioengineering and biotechnology* 7, p. 416.

Olughu, W., Deepika, G., Hewitt, C., and Rielly, C. (2019). "Insight into the large-scale upstream fermentation environment using scaled-down models". In: *Journal of Chemical Technology & Biotechnology* 94.3, pp. 647–657.

Olughu, W., Nienow, A., Hewitt, C., and Rielly, C. (2020). "Scale-down studies for the scale-up of a recombinant Corynebacterium glutamicum fed-batch fermentation: loss of homogeneity leads to lower levels of cadaverine production". In: *Journal of Chemical Technology & Biotechnology* 95.3, pp. 675–685.

Oosterhuis, N., Groesbeek, N., Olivier, A., and Kossen, N. (1983). "Scale-down aspects of the gluconic acid fermentation". In: *Biotechnology Letters* 5.3, pp. 141–146.

Oosterhuis, N. and Kossen, N. (1984). "Dissolved oxygen concentration profiles in a production-scale bioreactor". In: *Biotechnology and Bioengineering* 26.5, pp. 546–550.

Orth, J. D., Fleming, R. M., and Palsson, B. Ø. (2010). "Reconstruction and use of microbial metabolic networks: the core *Escherichia coli* metabolic model as an educational guide". In: *EcoSalPplus* 4.1.

Oyelade, J., Isewon, I., Oladipupo, F., Aromolaran, O., Uwoghiren, E., Ameh, F., Achas, M., and Adebiyi, E. (2016). "Clustering Algorithms: Their Application to Gene Expression Data". In: *Bioinformatics and biology insights* 10, pp. 237–253.

Pablos, T. E., Sigala, J. C., Le Borgne, S., and Lara, A. R. (2014). "Aerobic expression of Vitreoscilla hemoglobin efficiently reduces overflow metabolism in Escherichia coli". In: *Biotechnology journal* 9.6, pp. 791–799.

Papagianni, M. (2011). "Methodologies for scale-down of microbial bioprocesses". In: *J Microb Biochem Technol S* 5, p. 001.

Park, J. H. and Lee, S. Y. (2008). "Towards systems metabolic engineering of microorganisms for amino acid production". In: *Current Opinion in Biotechnology* 19.5, pp. 454–460.

Patel, V. C., Rodi, W., and Scheuerer, G. (1985). "Turbulence models for near-wall and low Reynolds number flows-A review". In: *AIAA Journal* 23.9, pp. 1308–1319.

Patil, V., Santos, C. N. S., Ajikumar, P. K., Sarria, S., and Takors, R. (2020). "Balancing glucose and oxygen uptake rates to enable high amorpha-4, 11-diene production in Escherichia coli via the methylerythritol phosphate (MEP) pathway". In: *Biotechnology and Bioengineering*.

Perez-Zabaleta, M., Sjöberg, G., Guevara-Martinez, M., Jarmander, J., Gustavsson, M., Quillaguaman, J., and Larsson, G. (2016). "Increasing the production of (R)-3-hydroxybutyrate in recombinant Escherichia coli by improved cofactor supply". In: *Microbial cell factories* 15.1, pp. 1–10.

Peters, J. M., Vangeloff, A. D., and Landick, R. (2011). "Bacterial Transcription Terminators: The RNA 3′-End Chronicles". In: *Journal of molecular biology* 412.5, pp. 793–813.

Pine, M. J. (1973). "Regulation of Intracellular Proteolysis in Escherichia coli". In: *Journal of bacteriology* 115.1, pp. 107–116.

Pirt, S. (1965). "The maintenance energy of bacteria in growing cultures". In: *Proceedings of the Royal Society of London. Series B. Biological Sciences* 163.991, pp. 224–231.

Posfai, G., Plunkett, G., Feher, T., Frisch, D., Keil, G. M., Umenhoffer, K., Kolisnychenko, V., Stahl, B., Sharma, S. S., De Arruda, M., et al. (2006). "Emergent properties of reduced-genome Escherichia coli". In: *science* 312.5776, pp. 1044–1046.

Proshkin, S., Rahmouni, A. R., Mironov, A., and Nudler, E. (2010). "Cooperation between translating ribosomes and RNA polymerase in transcription elongation". In: *Science* 328.5977, pp. 504–508.

Qian, H. and Elson, E. L. (2002). "Single-molecule enzymology: stochastic Michaelis–Menten kinetics". In: *Biophysical Chemistry* 101, pp. 565–576.

Reitzer, L. (2003). "Nitrogen assimilation and global regulation in Escherichia coli". In: *Annual Reviews in Microbiology* 57.1, pp. 155–176.

Reppas, N. B., Wade, J. T., Church, G. M., and Struhl, K. (2006). "The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting". In: *Molecular cell* 24.5, pp. 747–757.

Research, B. (2018). *In Report BIO030K - Global Markets for Enzymes in Industrial Applications: Global Markets.*

Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., Van Loo, P., Van Den Bossche, M., Catthoor, K., Sabbe, B., et al. (2012). "Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing". In: *Nature biotechnology* 30.1, pp. 61–68.

Reynolds, O. (1895). "IV. On the dynamical theory of incompressible viscous fluids and the determination of the criterion". In: *Phil. Trans. R. Soc. Lond. A* 186, pp. 123–164.

Rhodius, V. A. and Mutalik, V. K. (2010). "Predicting strength and function for promoters of the Escherichia coli alternative sigma factor, sigmaE". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.7, pp. 2854–2859.

Rizzi, M., Baltes, M., Theobald, U., and Reuss, M. (1997). "In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae*: II. mathematical model". In: *Biotechnology and Bioengineering* 55.4, pp. 592–608.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data". In: *Bioinformatics* 26.1, pp. 139–140.

Robinson, M. D. and Oshlack, A. (2010). "A scaling normalization method for differential expression analysis of RNA-seq data". In: *Genome Biology* 11.3, R25.

Rodi, W. (2017). "Turbulence Modeling and Simulation in Hydraulics: A Historical Review". In: *Journal of Hydraulic Engineering* 143.5.

Rousseeuw, P. J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65.

Sajjadi, B., Raman, A. A. A., Shah, R. S. S. R. E., and Ibrahim, S. (2013). "Review on applicable breakup/coalescence models in turbulent liquid-liquid flows". In: *Reviews in Chemical Engineering* 29.3, pp. 131–158.

Salgado, H., Gama-Castro, S., Martinez-Antonio, A., Diaz-Peredo, E., Sanchez-Solano, F., Peralta-Gil, M., Garcia-Alonso, D., Jimenez-Jacinto, V., Santos-Zavaleta, A., Bonavides-Martinez, C., et al. (2004). "RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12". In: *Nucleic Acids Research* 32.suppl 1, pp. D303–D306.

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sanchez-Perez, M., Gomez-Romero, L., Ledezma-Tejeida, D., Garcia-Sotelo, J. S., Alquicira-Hernandez, K., Muniz-Rascado, L. J., Pena-Loredo, P., et al. (2019). "RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12". In: *Nucleic acids research* 47.D1, pp. D212–D220.

Sarkizi Shams Hajian, C., Haringa, C., Noorman, H., and Takors, R. (2020). "Predicting By-Product Gradients of Baker's Yeast Production at Industrial Scale: A Practical Simulation Approach". In: *Processes* 8.12, p. 1554.

Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., and Schomburg, D. (2010). "BRENDA, the enzyme information system in 2011". In: *Nucleic Acids Research* 39.suppl_1, pp. D670–D676.

Schlitt, T. and Brazma, A. (2007). "Current approaches to gene regulatory network modelling". In: *BMC bioinformatics* 8.Suppl 6, S9.

Schmalzriedt, S., Jenne, M., Mauch, K., and Reuss, M. (2003). "Integration of physiology and fluid dynamics". In: *Process Integration in Biochemical Engineering*. Springer, pp. 19–68.

Schmid, A., Kollmer, A., Mathys, R. G., and Witholt, B. (1998). "Developments toward large-scale bacterial bioprocesses in the presence of bulk amounts of organic solvents". In: *Extremophiles* 2.3, pp. 249–256.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). "The RIN: an RNA integrity number for assigning integrity values to RNA measurements". In: *BMC molecular biology* 7.1, pp. 1–14.

Shamir, M., Bar-On, Y., Phillips, R., and Milo, R. (2016). "Snapshot: timescales in cell biology". In: *Cell* 164.6, pp. 1302–1302.

Sharma, A. K., Shukla, E., Janoti, D. S., Mukherjee, K. J., and Shiloach, J. (2020). "A novel knock out strategy to enhance recombinant protein expression in Escherichia coli". In: *Microbial cell factories* 19.1, pp. 1–10.

Sharma, S. S., Blattner, F. R., and Harcum, S. W. (2007). "Recombinant protein production in an Escherichia coli reduced genome strain". In: *Metabolic engineering* 9.2, pp. 133–141.

Sharma, S. S., Campbell, J. W., Frisch, D., Blattner, F. R., and Harcum, S. W. (2007). "Expression of two recombinant chloramphenicol acetyltransferase variants in highly reduced genome Escherichia coli strains". In: *Biotechnology and bioengineering* 98.5, pp. 1056–1070.

Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., and Liao, J. C. (2011). "High titer anaerobic 1-butanol synthesis in Escherichia coli enabled by driving forces". In: *Applied and environmental microbiology*.

Shendure, J. and Ji, H. (2008). "Next-generation DNA sequencing". In: *Nature biotechnology* 26.10, pp. 1135–1145.

Shih, T.-H., Liou, W. W., Shabbir, A., Yang, Z., and Zhu, J. (1994). "A new k-epsilon eddy viscosity model for high Reynolds number turbulent flows: Model development and validation". In: *NASA Sti/recon Technical Report N* 95, p. 11442.

Shimada, T., Fujita, N., Yamamoto, K., and Ishihama, A. (2011). "Novel roles of cAMP receptor protein (CRP) in regulation of transport and metabolism of carbon sources". In: *PLoS One* 6.6.

Shmulevich, I., Dougherty, E. R., Kim, S., and Zhang, W. (2002). "Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks". In: *Bioinformatics* 18.2, pp. 261–274.

Silva, I. J., Saramago, M., Dressaire, C., Domingues, S., Viegas, S. C., and Arraiano, C. M. (2011). "Importance and key events of prokaryotic RNA decay: the ultimate fate of an RNA molecule". In: *Wiley interdisciplinary reviews. RNA* 2.6, pp. 818–836.

Simen, J. D., Löffler, M., Jäger, G., Schäferhoff, K., Freund, A., Matthes, J., Müller, J., and Takors, R. (2017a). "Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations". In: *Microbial Biotechnology*, 10: 858–872.

Simen, J. D., Löffler, M., Jäger, G., Schäferhoff, K., Freund, A., Matthes, J., Müller, J., and Takors, R. (2017b). "Transcriptional response of Escherichia coli to ammonia and glucose fluctuations". In: *Microbial biotechnology* 10.4, pp. 858–872.

Sommerfeld, M. and Decker, S. (2004). "State of the art and future trends in CFD simulation of stirred vessel hydrodynamics". In: *Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology* 27.3, pp. 215–224.

Storz, G., Opdyke, J. A., and Zhang, A. (2004). "Controlling mRNA stability and translation with small, noncoding RNAs". In: *Current Opinion in Microbiology* 7.2, pp. 140–144.

Stouthamer, A. (1973). "A theoretical study on the amount of ATP required for synthesis of microbial cell material". In: *Antonie van Leeuwenhoek* 39.1, pp. 545–565.

Stouthamer, A. H. and Bettenhaussen, C. (1973). "Utilization of energy for growth and maintenance in continuous and batch cultures of microorganisms: A reevaluation of the method for the determination of ATP production by measuring molar growth yields". In: *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics* 301.1, pp. 53–70.

Subramaniam, A. R., Zid, B. M., and O'Shea, E. K. (2014). "An Integrated Approach Reveals Regulatory Controls on Bacterial Translation Elongation". In: *Proceedings of the National Academy of Sciences* 159.5, pp. 1200–1211.

Sun, C., Guo, Y., Tang, K., Wen, Z., Li, B., Zeng, Z., and Wang, X. (2017). "MqsR/MqsA toxin/antitoxin system regulates persistence and biofilm formation in Pseudomonas putida KT2440". In: *Frontiers in microbiology* 8, p. 840.

Sun, X., Zahir, Z., Lynch, K. H., and Dennis, J. J. (2011). "An antirepressor, SrpR, is involved in transcriptional regulation of the SrpABC solvent tolerance efflux pump of Pseudomonas putida S12". In: *Journal of bacteriology* 193.11, pp. 2717–2725.

Sundararaman, N., Ash, C., Guo, W., Button, R., Singh, J., and Feng, X. (2015). "iTAP: integrated transcriptomics and phenotype database for stress response of *Escherichia coli* and *Saccharomyces cerevisiae*". In: *BMC Research Notes* 8.1, p. 771.

Sunya, S., Gorret, N., Delvigne, F., Uribelarrea, J.-L., and Molina-Jouve, C. (2012). "Real-time monitoring of metabolic shift and transcriptional induction of yciG:: luxCDABE E. coli reporter strain to a glucose pulse of different concentrations". In: *Journal of biotechnology* 157.3, pp. 379–390.

Szenk, M., Dill, K. A., and Graff, A. M. de (2017). "Why do fast-growing bacteria enter overflow metabolism? Testing the membrane real estate hypothesis". In: *Cell Systems* 5.2, pp. 95–104.

Tabib, M., Siddiqui, M. S., Rasheed, A., and Kvamsdal, T. (2017). "Industrial scale turbine and associated wake development-comparison of RANS based Actuator Line Vs Sliding Mesh Interface Vs Multiple Reference Frame method." In: *Energy Procedia* 137, pp. 487–496.

Taheri-Araghi, S. and Jun, S. (2016). "Single-Cell Cultivation in Microfluidic Devices". In: *Hydrocarbon and Lipid Microbiology Protocols: Single-Cell and Single-Molecule Methods*, pp. 5–16.

Takors, R. (2012). "Scale-up of microbial processes: impacts, tools and open questions". In: *Journal of Biotechnology* 160.1, pp. 3–9.

Tang, W., Deshmukh, A. T., Haringa, C., Wang, G., Gulik, W. van, Winden, W. van, Reuss, M., Heijnen, J. J., Xia, J., Chu, J., et al. (2017). "A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by *Penicillium chrysogenum*". In: *Biotechnology and Bioengineering*, 114: 1733–1743.

Tang, X., Tan, Y., Zhu, H., Zhao, K., and Shen, W. (2009). "Microbial conversion of glycerol to 1, 3-propanediol by an engineered strain of Escherichia coli". In: *Applied and environmental microbiology* 75.6, pp. 1628–1634.

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., Krylov, D. M., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., et al. (2003). "The

COG database: an updated version includes eukaryotes". In: *BMC bioinformatics* 4.1, pp. 1–14.

Taymaz-Nikerel, H., Borujeni, A. E., Verheijen, P. J., Heijnen, J. J., and Gulik, W. M. van (2010). "Genome-derived minimal metabolic models for Escherichia coli MG1655 with estimated in vivo respiratory ATP stoichiometry". In: *Biotechnology and bioengineering* 107.2, pp. 369–381.

Taymaz-Nikerel, H., Van Gulik, W. M., and Heijnen, J. J. (2011). "Escherichia coli responds with a rapid and large change in growth rate upon a shift from glucose-limited to glucose-excess conditions". In: *Metabolic engineering* 13.3, pp. 307–318.

Teleki, A., Sánchez-Kopper, A., and Takors, R. (2015). "Alkaline conditions in hydrophilic interaction liquid chromatography for intracellular metabolite quantification using tandem mass spectrometry". In: *Analytical Biochemistry* 475, pp. 4–13.

Theisen, M. and Liao, J. C. (2017). "Industrial biotechnology: Escherichia coli as a host". In: *Industrial Biotechnology: Microorganisms* 1, pp. 149–181.

Theresa Scharl and Friedrich Leisch (2006). "The stochastic QT-clust algorithm: evaluation of stability and variance on time-course microarray data". In: *In Compstat 2006 – Proceedings in Computational Statistics*. Ed. by V. M. Rizzi A. Physica Verlag, pp. 1015–1022.

Thirumangalathu, R., Krishnan, S., Ricci, M. S., Brems, D. N., Randolph, T. W., and Carpenter, J. F. (2009). "Silicone oil-and agitation-induced aggregation of a monoclonal antibody in aqueous solution". In: *Journal of pharmaceutical sciences* 98.9, pp. 3167–3181.

Thomas, R. (1973). "Boolean formalization of genetic control circuits". In: *Journal of Theoretical Biology* 42.3, pp. 563–585.

Tolić-Nørrelykke, S. F., Engh, A. M., Landick, R., and Gelles, J. (2004). "Diversity in the rates of transcript elongation by single RNA polymerase molecules". In: *The Journal of biological chemistry* 279.5, pp. 3292–3299.

Tomasek, K., Bergmiller, T., and Guet, C. C. (2018). "Lack of cations in flow cytometry buffers affect fluorescence signals by reducing membrane stability and viability of Escherichia coli strains". In: *Journal of biotechnology* 268, pp. 40–52.

Traxler, M. F., Summers, S. M., Nguyen, H.-T., Zacharia, V. M., Hightower, G. A., Smith, J. T., and Conway, T. (2008). "The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli". In: *Molecular microbiology* 68.5, pp. 1128–1148.

Traxler, M. F., Zacharia, V. M., Marquardt, S., Summers, S. M., Nguyen, H.-T., Stark, S. E., and Conway, T. (2011). "Discretely calibrated regulatory loops controlled by ppGpp

partition gene induction across the 'feast to famine'gradient in *Escherichia coli*". In: *Molecular Microbiology* 79.4, pp. 830–845.

Tsukagoshi, N. and Aono, R. (2000). "Entry into and Release of Solvents byEscherichia coli in an Organic-Aqueous Two-Liquid-Phase System and Substrate Specificity of the AcrAB-TolC Solvent-Extruding Pump". In: *Journal of Bacteriology* 182.17, pp. 4803–4810.

Turner, T. E., Schnell, S., and Burrage, K. (2004). "Stochastic approaches for modelling in vivo reactions". In: *Computational Biology and Chemistry* 28.3, pp. 165–178.

Valgepea, K., Peebo, K., Adamberg, K., and Vilu, R. (2015). "Lean-proteome strains–next step in metabolic engineering". In: *Frontiers in bioengineering and biotechnology* 3, p. 11.

Vallino, J. J. and Stephanopoulos, G. (1993). "Metabolic flux distributions in Corynebacterium glutamicum during growth and lysine overproduction". In: *Biotechnology and Bioengineering* 41.6, pp. 633–646.

van der Laan, Mark J. and Pollard, K. S. (2003). "A new algorithm for hybrid hierarchical clustering with visualization and the bootstrap". In: *Journal of Statistical Planning and Inference* 117.2, pp. 275–303.

Van Dijk, E. L., Jaszczyszyn, Y., and Thermes, C. (2014). "Library preparation methods for next-generation sequencing: tone down the bias". In: *Experimental cell research* 322.1, pp. 12–20.

Van Ede, C., Van Houten, R., and Beenackers, A. (1995). "Enhancement of gas to water mass transfer rates by a dispersed organic phase". In: *Chemical engineering science* 50.18, pp. 2911–2922.

Van Vliet, A. H. (2010). "Next generation sequencing of microbial transcriptomes: challenges and opportunities". In: *FEMS microbiology letters* 302.1, pp. 1–7.

Van't Riet, K. (1979). "Review of measuring methods and results in nonviscous gas-liquid mass transfer in stirred vessels". In: *Industrial & Engineering Chemistry Process Design and Development* 18.3, pp. 357–364.

Vanderpool, C. K. and Gottesman, S. (2004). "Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system". In: *Molecular microbiology* 54.4, pp. 1076–1089.

Vasilakou, E., Van Loosdrecht, M. C., and Wahl, S. A. (2020). "Escherichia coli metabolism under short-term repetitive substrate dynamics: adaptation and trade-offs". In: *Microbial Cell Factories* 19, pp. 1–19.

Veit, A., Polen, T., and Wendisch, V. F. (2007). "Global gene expression analysis of glucose overflow metabolism in Escherichia coli and reduction of aerobic acetate formation". In: *Applied microbiology and biotechnology* 74.2, pp. 406–421.

Venneker, B. C., Derksen, J. J., and Van den Akker, H. E. (2002). "Population balance modeling of aerated stirred vessels based on CFD". In: *AIChE Journal* 48.4, pp. 673–685.

Versteeg, H. K. and Malalasekera, W. (2007). *An introduction to computational fluid dynamics: the finite volume method*. Pearson Education.

Viegas, S. C., Silva, I. J., Saramago, M., Domingues, S., and Arraiano, C. M. (2011). "Regulation of the small regulatory RNA MicA by ribonuclease III: a target-dependent pathway". In: *Nucleic acids research* 39.7, pp. 2918–2930.

Villadsen, J., Nielsen, J., and Lidén, G. (2011). *Bioreaction engineering principles*. Springer Science & Business Media.

Vrabel, P., Lans, R. G. van der, Luyben, K. C. A., Boon, L., and Nienow, A. W. (2000). "Mixing in large-scale vessels stirred with multiple radial or radial and axial up-pumping impellers: modelling and measurements". In: *Chemical Engineering Science* 55.23, pp. 5881–5896.

Wang, G., Haringa, C., Tang, W., Noorman, H., Chu, J., Zhuang, Y., and Zhang, S. (2020). "Coupled metabolic-hydrodynamic modeling enabling rational scale-up of industrial bioprocesses". In: *Biotechnology and Bioengineering* 117.3, pp. 844–867.

Wang, J. D. and Levin, P. A. (2009). "Metabolism, cell growth and the bacterial cell cycle". In: *Nature Reviews Microbiology* 7.11, pp. 822–827.

Wang, Q., Yu, H., Xia, Y., Kang, Z., and Qi, Q. (2009). "Complete PHB mobilization in Escherichia coli enhances the stress tolerance: a potential biotechnological application". In: *Microbial cell factories* 8.1, pp. 1–9.

Wang, R.-S., Wang, Y., Zhang, X.-S., and Chen, L. (2007). "Inferring transcriptional regulatory networks from high-throughput data". In: *Bioinformatics* 23.22, pp. 3056–3064.

Wang, T., Wang, J., and Jin, Y. (2005). "Population balance model for gas- liquid flows: Influence of bubble coalescence and breakup models". In: *Industrial & Engineering Chemistry Research* 44.19, pp. 7540–7549.

Wang, Z., Gerstein, M., and Snyder, M. (2009). "RNA-Seq: a revolutionary tool for transcriptomics". In: *Nature Reviews Genetics* 10.1, pp. 57–63.

Ward, V. C., Chatzivasileiou, A. O., and Stephanopoulos, G. (2018). "Metabolic engineering of Escherichia coli for the production of isoprenoids". In: *FEMS microbiology letters* 365.10, p. 79.

Watson, J. D. and Baker, T. A. (2014). *Molecular biology of the gene*.

Weber, J., Hoffmann, F., and Rinas, U. (2002). "Metabolic adaptation of Escherichia coli during temperature-induced recombinant protein production: 2. Redirection of metabolic fluxes". In: *Biotechnology and bioengineering* 80.3, pp. 320–330.

Wegerer, A., Sun, T., and Altenbuchner, J. (2008). "Optimization of an E. coli L-rhamnose-inducible expression vector: test of various genetic module combinations". In: *BMC biotechnology* 8.1, pp. 1–12.

Wehrs, M., Tanjore, D., Eng, T., Lievense, J., Pray, T. R., and Mukhopadhyay, A. (2019). "Engineering robust production microbes for large-scale cultivation". In: *Trends in microbiology* 27.6, pp. 524–537.

Wen, Q., Chen, Z., Tian, T., and Chen, W. (2010). "Effects of phosphorus and nitrogen limitation on PHA production in activated sludge". In: *Journal of environmental sciences* 22.10, pp. 1602–1607.

Werner, S., Kaiser, S. C., Kraume, M., and Eibl, D. (2014). "Computational fluid dynamics as a modern tool for engineering characterization of bioreactors". In: *Pharmaceutical Bioprocessing* 2.1, pp. 85–99.

Westerhoff, H. V. and Palsson, B. O. (2004). "The evolution of molecular biology into systems biology". In: *Nature Biotechnology* 22.10, pp. 1249–1252.

Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). "Unicycler: resolving bacterial genome assemblies from short and long sequencing reads". In: *PLoS computational biology* 13.6.

Wilkinson, D. J. (2011). *Stochastic modelling for systems biology*. CRC press.

Williamson, G., Tamburrino, G., Bizior, A., Boeckstaens, M., Mirandela, G. D., Bage, M. G., Pisliakov, A., Ives, C. M., Terras, E., Hoskisson, P. A., et al. (2020). "A two-lane mechanism for selective biological ammonium transport". In: *Elife* 9.

Wilms, B., Hauck, A., Reuss, M., Syldatk, C., Mattes, R., Siemann, M., and Altenbuchner, J. (2001). "High-cell-density fermentation for production of L-N-carbamoylase using an expression system based on the Escherichia coli rhaBAD promoter". In: *Biotechnology and bioengineering* 73.2, pp. 95–103.

Wintermeyer, W., Peske, F., Beringer, M., Gromadski, K. B., Savelsbergh, A., and Rodnina, M. V. (2004). "Mechanisms of elongation on the ribosome: dynamics of a macromolecular machine". In: *Biochemical Society transactions* 32.Pt 5, pp. 733–737.

Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., et al. (2011). "SABIO-RK—database for biochemical reaction kinetics". In: *Nucleic Acids Research* 40.D1, pp. D790–D796.

Wittmann, C., Liao, J., Lee, S., Nielsen, J., and Stephanopoulos, G. (2016). *Industrial Biotechnology: Microorganisms*. Wiley.

Wittmann, C. and Lee, S. Y. (2012). *Systems metabolic engineering*. Springer Science & Business Media.

Wubbolts, M. G., Favre-Bulle, O., and Witholt, B. (1996). "Biosynthesis of synthons in two-liquid-phase media". In: *Biotechnology and bioengineering* 52.2, pp. 301–308.

Wulffen, J. von, Buchholz, P., Sawodny, O., and Feuer, R. (2015). "Modeling the metabolism of *Escherichia coli* under oxygen gradients with dynamically changing flux bounds". In: *Bioinformatics and Bioengineering, 2015*, pp. 1–6.

Wulffen, J. von, RecogNice-Team, Sawodny, O., and Feuer, R. (2016). "Transition of an anaerobic Escherichia coli culture to aerobiosis: balancing mRNA and protein levels in a demand-directed dynamic flux balance analysis". In: *PLoS One* 11.7.

Wulffen, J. von, Ulmer, A., Jäger, G., Sawodny, O., and Feuer, R. (2017). "Rapid Sampling of *Escherichia coli* After Changing Oxygen Conditions Reveals Transcriptional Dynamics". In: *Genes* 8.3, p. 90.

Wutz, J., Lapin, A., Siebler, F., Schäfer, J. E., Wucherpfennig, T., Berger, M., and Takors, R. (2016). "Predictability of kLa in stirred tank reactors under multiple operating conditions using an Euler–Lagrange approach". In: *Engineering in Life Sciences* 16.7, pp. 633–642.

Xi, H., Schneider, B. L., and Reitzer, L. (2000). "Purine catabolism in Escherichia coliand function of xanthine dehydrogenase in purine salvage". In: *Journal of bacteriology* 182.19, pp. 5332–5341.

Xu, B., Jahic, M., Blomsten, G., and Enfors, S.-O. (1999). "Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*". In: *Applied Microbiology and Biotechnology* 51.5, pp. 564–571.

Xu, Q., Nakajima, M., Ichikawa, S., Nakamura, N., Roy, P., Okadome, H., and Shiina, T. (2009). "Effects of surfactant and electrolyte concentrations on bubble formation and stabilization". In: *Journal of Colloid and Interface science* 332.1, pp. 208–214.

Yeoh, G., Cheung, C., and Tu, J. (2014). "Chapter 3-Population Balance Approach—A Generic Framework". In: *Multiphase Flow Analysis Using Population Balance Modeling*, pp. 69–90.

Yeung, K. Y., Haynor, D. R., and Ruzzo, W. L. (2001). "Validating clustering for gene expression data". In: *Bioinformatics (Oxford, England)* 17.4, pp. 309–318.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., et al. (2011). "Metabolic engineering of

Escherichia coli for direct production of 1, 4-butanediol". In: *Nature chemical biology* 7.7, pp. 445–452.

Yoshida, F., Yamane, T., and Miyamoto, Y. (1970). "Oxygen absorption into oil-in-water emulsions. A study on hydrocarbon fermentors". In: *Industrial & Engineering Chemistry Process Design and Development* 9.4, pp. 570–577.

Zhao, J., Li, Q., Sun, T., Zhu, X., Xu, H., Tang, J., Zhang, X., and Ma, Y. (2013). "Engineering central metabolic modules of Escherichia coli for improving β-carotene production". In: *Metabolic engineering* 17, pp. 42–50.

Zhao, Y. and Karypis, G. (2005). "Data Clustering in Life Sciences". In: *Molecular Biotechnology* 31.1, pp. 055–080.

Zhou, J., Liu, L., Shi, Z., Du, G., and Chen, J. (2009). "ATP in current biotechnology: regulation, applications and perspectives". In: *Biotechnology advances* 27.1, pp. 94–101.

Ziegler, M., Zieringer, J., and Takors, R. (2020). "Transcriptional profiling of the stringent response mutant strain E. coli SR reveals enhanced robustness to large-scale conditions". In: *Microbial Biotechnology*.

Zieringer, J. and Takors, R. (2018). "In silico prediction of large-scale microbial production performance: constraints for getting proper data-driven models". In: *Computational and structural biotechnology journal* 16, pp. 246–256.

Zieringer, J., Wild, M., and Takors, R. (2021). "Data-driven in silico prediction of regulation heterogeneity and ATP demands of Escherichia coli in large-scale bioreactors". In: *Biotechnology and Bioengineering* 118.1, pp. 265–278.

Zlokarnik, M. (2013). *Rührtechnik: Theorie und Praxis*. Springer-Verlag.